

Robust estimators of accelerated failure time regression with generalized log-gamma errors

Claudio Agostinelli¹, Isabella Locatelli², Alfio Marazzi^{*2}, and
Víctor J. Yohai³

¹*Department of Mathematics, University of Trento, Trento, Italy*

²*Institute of social and preventive medicine, Lausanne University Hospital,
Switzerland*

³*Departamento de Matemáticas, Facultad de Ciencias Exactas y Naturales,
University of Buenos Aires and CONICET*

December 7, 2015

Abstract

The generalized log-gamma (GLG) model is a very flexible family of distributions to analyze datasets in many different areas of science and technology. In this paper, we propose estimators which are simultaneously highly robust and highly efficient for the parameters of a GLG distribution in the presence of censoring. We also introduced estimators with the same properties for accelerated failure time models with censored observations and error distribution belonging to the GLG family. We prove that the proposed estimators are asymptotically fully efficient and examine the maximum mean square error using Monte Carlo simulations. The simulations confirm that the proposed estimators are highly robust and highly efficient for finite sample size. Finally, we illustrate the good behavior of the proposed estimators with two real datasets.

Keywords: Censored data. Quantile distance estimates. τ estimators. Truncated maximum likelihood estimators. Weighted likelihood estimators.

^{*}Corresponding author alfio.marazzi@chuv.ch

1 Introduction

Generalized log-gamma (GLG) regression with censored observations is a large subclass of Accelerated Failure Time (AFT) models introduced by Lawless (1980). Many models broadly used in the lifetime data analysis – including lognormal, log-gamma, and log-Weibull regression – are specific cases of GLG regression. GLG regression has been widely applied in various areas of survival analysis (e.g. Kim et al., 1993; Sun et al., 1999; Abadi et al., 2012).

Usually, the parameters are estimated by means of the maximum likelihood (ML) principle, which provides fully efficient estimators when the observations follow the model. Unfortunately, the ML estimator is extremely sensitive to the presence of outliers in the sample.

There are several proposals of diagnostic tools to detect outliers and assess their influence on the GLG regression parameter estimates (e.g. Ortega et al., 2003, 2008; Silva et al., 2010). Moreover, two families of robust estimators of the GLG model without censored observations and without covariate information have been introduced by Agostinelli et al. (2014a) for models with three parameters: location, scale, and shape. These families of estimators are: the (weighted) quantile τ ($Q\tau$) estimators and the one-step weighted likelihood (1SWL) estimators. A $Q\tau$ estimator minimizes a τ scale (Yohai and Zamar, 1988) of the differences between empirical and theoretical quantiles. It is $n^{1/2}$ consistent but not asymptotically normal. However, it is a convenient starting point to define the 1SWL-estimator.

In this paper, we extend the $Q\tau$ estimator proposed in Agostinelli et al. (2014a) to GLG regression with right censoring by introducing the trimmed $Q\tau$ -estimator (T $Q\tau$ -estimator); we also extend the truncated maximum likelihood (TML) estimator introduced in Marazzi and Yohai (2004) to GLG regression and. To improve the robustness of this estimator without modifying its asymptotic efficiency we also define a one-step version of the TML estimator (1TML-estimator). For the sake of completeness, we also define an extension of the 1SWL estimator which is fully described in the Appendix. However a Monte Carlo study show that this estimator is much less robust than the T $Q\tau$ - and 1TML- estimators.

The procedures introduced here for the GLG family can be applied to other location-scale-shape models, such as the three-parameter log-Weibull family.

Section 2 defines the $Q\tau$ - and T $Q\tau$ -estimators for censored observations in the absence of covariates. Section 3 describes the TML estimators. Section 4

extends the estimators to the regression case. Section 5 shows the results of a Monte Carlo study comparing the performance of the proposed methods for finite sample sizes. Section 6 discusses two examples with real data. Section 7 provides concluding remarks.

Section A is an Appendix that include the proofs and an extension for censored data of the one step weighted likelihood estimator used in Agostinelli et al. (2014a). For completeness sake, this estimator is included in our Monte Carlo study.

2 The Q_{τ} - and TQ_{τ} -estimators for censored observations without covariates

2.1 The generalized log-gamma distribution

The GLG family of distributions depends on three parameters μ , σ , and λ . We use the parametrization of Prentice (1974) and denote the family by $GLG(\mu, \sigma, \lambda)$, $\mu \in \mathbb{R}$, $\sigma > 0$, $\lambda \in \mathbb{R}$. A random variable y has a $GLG(\mu, \sigma, \lambda)$ distribution if

$$y = \mu + \sigma u \quad (1)$$

and u has density

$$f_{\lambda}(u) = \begin{cases} \frac{|\lambda|}{\Gamma(\lambda^{-2})} (\lambda^{-2})^{\lambda^{-2}} \exp((\lambda^{-2})(\lambda u - e^{\lambda u})) & \text{if } \lambda \neq 0, \\ \frac{1}{\sqrt{2\pi}} \exp(-\frac{u^2}{2}) & \text{if } \lambda = 0, \end{cases} \quad (2)$$

where Γ denotes the Gamma function. This family includes many common models, such as the log-Weibull model ($\lambda = 1$), the log-exponential model ($\lambda = 1$ and $\sigma = 1$), the log-gamma model ($\sigma = \lambda$), and the normal model ($\lambda = 0$). The density of y is

$$f_{\boldsymbol{\theta}}(y) = \frac{1}{\sigma} f_{\lambda} \left(\frac{y - \mu}{\sigma} \right), \quad (3)$$

where $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$.

Suppose that y_1, \dots, y_n are n i.i.d. random times with a GLG cdf $F_{\boldsymbol{\theta}_0}(y)$. We want to estimate $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$. We consider single censoring on the right, i.e., the true value of y_i is not observed. Instead, the censored variable $y_i^* = \min(y_i, c_i)$ is observed, where c_1, \dots, c_n are i.i.d. censoring times, which

are independent of the y_i 's. We define the censoring indicator $\delta_i = 1$ if $y_i^* = y_i$ and $\delta_i = 0$ if $y_i^* = c_i$. Let $\mathbf{z}_i = (y_i^*, \delta_i)$ and G_n be the empirical distribution function based on $(\mathbf{z}_1, \dots, \mathbf{z}_n)$.

2.2 Score functions and ML estimator

The ML estimator of the parameters of a GLG model under censoring can be easily defined as follows. Let $S_{\boldsymbol{\theta}}(y) = 1 - F_{\boldsymbol{\theta}}(y) = 1 - F_{\lambda}((y - \mu)/\sigma)$ denote the survival function. Then, the negative log-likelihood function is

$$- \sum_{i=1}^n [\delta_i \log f_{\boldsymbol{\theta}}(y_i^*) + (1 - \delta_i) \log S_{\boldsymbol{\theta}}(y_i^*)]. \quad (4)$$

In the absence of censoring, the score functions $\mathbf{d} = (d_1, d_2, d_3)^\top$ are:

$$d_1(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \mu} \log f_{\boldsymbol{\theta}}(y) = \frac{1}{\sigma} \xi_{\lambda}(u), \quad (5)$$

$$d_2(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \sigma} \log f_{\boldsymbol{\theta}}(y) = \frac{1}{\sigma} (\xi_{\lambda}(u)u + 1), \quad (6)$$

$$d_3(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \lambda} \log f_{\boldsymbol{\theta}}(y) = \psi_{\lambda}(u), \quad (7)$$

where $u = (y - \mu)/\sigma$, $\dot{F}_{\boldsymbol{\theta}}(y) = \dot{F}_{\lambda}(u) = dF_{\lambda}(u)/d\lambda$,

$$\xi_{\lambda}(u) = \frac{f'_{\lambda}(u)}{f_{\lambda}(u)} = \frac{(1 - e^{\lambda u})}{\lambda},$$

$$\psi_{\lambda}(u) = -\frac{\partial}{\partial \lambda} \log f_{\lambda}(u) = \frac{1}{\lambda^3} (2\zeta(\lambda) - \lambda^2 + \lambda u - \exp(\lambda u)(2 - \lambda u)),$$

$\zeta(\lambda) = -2 \log(\lambda) - \dot{\Gamma}(\lambda^{-2}) + 1$, and $\dot{\Gamma}$ denotes the digamma function. Let

$$s_1(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \mu} \log S_{\boldsymbol{\theta}}(y) = -\frac{1}{\sigma} \frac{f_{\lambda}(u)}{S_{\lambda}(u)}, \quad (8)$$

$$s_2(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \sigma} \log S_{\boldsymbol{\theta}}(y) = -\frac{1}{\sigma} \frac{f_{\lambda}(u)u}{S_{\lambda}(u)} = s_1(y, \boldsymbol{\theta})u, \quad (9)$$

$$s_3(y, \boldsymbol{\theta}) = -\frac{\partial}{\partial \lambda} \log S_{\boldsymbol{\theta}}(y) = \frac{\dot{F}_{\lambda}(u)}{S_{\lambda}(u)}. \quad (10)$$

Then, the score functions for the case with censored observations are

$$v_k(y, \delta, \boldsymbol{\theta}) = \delta d_k(y, \boldsymbol{\theta}) + (1 - \delta) s_k(y, \boldsymbol{\theta}), \quad k = 1, 2, 3. \quad (11)$$

The *ML estimator* of $\boldsymbol{\theta}$ is given by the following system of equations

$$E_{G_n}(\mathbf{v}(y, \delta, \boldsymbol{\theta})) = \frac{1}{n} \sum_{i=1}^n \mathbf{v}(y_i^*, \delta_i, \boldsymbol{\theta}) = \mathbf{0}, \quad (12)$$

where $\mathbf{v} = (v_1, v_2, v_3)^\top$ is the score function vector. It is easy to show that

$$\nabla_{\boldsymbol{\theta}} S_{\boldsymbol{\theta}}(y^*) = E_{\boldsymbol{\theta}}(\mathbf{d}(y, \boldsymbol{\theta}) | y > y^*),$$

where $\nabla_{\boldsymbol{\theta}}$ indicates differentiation w.r.t. $\boldsymbol{\theta}$. Hence, an alternative expression for the likelihood equations is

$$\frac{1}{n} \sum_{i=1}^n \delta_i \mathbf{d}(y_i^*, \boldsymbol{\theta}) + (1 - \delta_i) E_{\boldsymbol{\theta}}(\mathbf{d}(y, \boldsymbol{\theta}) | y > y_i^*) = \mathbf{0}.$$

Following Locatelli et al. (2010), we define the *semiempirical cdf* of y for a given $\boldsymbol{\theta}$ as

$$H_{n,\boldsymbol{\theta}}(z) = \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}}[I(y \leq z) | y_i^*, \delta_i]$$

or equivalently,

$$H_{n,\boldsymbol{\theta}}(z) = \frac{1}{n} \sum_{i=1}^n \delta_i I(y_i \leq z) + \frac{1}{n} \sum_{i=1}^n (1 - \delta_i) \frac{[F_{\boldsymbol{\theta}}(z) - F_{\boldsymbol{\theta}}(y_i)]^+}{1 - F_{\boldsymbol{\theta}}(y_i)}. \quad (13)$$

Thus, when there is no censoring, $H_{n,\boldsymbol{\theta}}(z)$ coincides with the usual empirical cdf. If $\tilde{\boldsymbol{\theta}}$ is a consistent estimator of $\boldsymbol{\theta}_0$, then $H_{n,\tilde{\boldsymbol{\theta}}}(z)$ is a consistent estimator of $F_{\boldsymbol{\theta}_0}(z)$ and, for any measurable function $h(y)$, we have $\lim_{n \rightarrow \infty} E_{n,\tilde{\boldsymbol{\theta}}}[h(y)] = E_{\boldsymbol{\theta}_0}[h(y)]$ a.s., where $E_{n,\boldsymbol{\theta}}$ denotes expectation under $H_{n,\boldsymbol{\theta}}$. Finally, another expression of the likelihood equations is

$$E_{n,\boldsymbol{\theta}}(\mathbf{d}(y, \boldsymbol{\theta})) = \mathbf{0}. \quad (14)$$

Let $\mathbf{M}(\boldsymbol{\theta}) = E(\mathbf{v}(y, \delta, \boldsymbol{\theta})\mathbf{v}(y, \delta, \boldsymbol{\theta})^\top)$ and $\mathbf{G}(\boldsymbol{\theta}) = E(\nabla_{\boldsymbol{\theta}} \mathbf{v}(y, \delta, \boldsymbol{\theta}))$ then the asymptotic covariance matrix of the ML estimator is

$$\boldsymbol{\Sigma}(\boldsymbol{\theta}) = \mathbf{G}(\boldsymbol{\theta})^{-1} \mathbf{M}(\boldsymbol{\theta}) \mathbf{G}(\boldsymbol{\theta})^{-\top}.$$

2.3 The trimmed $Q\tau$ estimator

Agostinelli et al. (2014a) define the quantile τ ($Q\tau$) estimator and the weighted $Q\tau$ ($WQ\tau$) estimator for non-censored i.i.d. observations as follows. For $0 < u < 1$ let $Q(u, \boldsymbol{\theta})$ denote the u -quantile of $F_{\boldsymbol{\theta}}(y)$. Then, $Q(u, \boldsymbol{\theta}) = \sigma Q^*(u, \lambda) + \mu$, where $Q^*(u, \lambda) = Q(u, (0, 1, \lambda))$. Given a sample y_1, \dots, y_n , let F_n denote the empirical cdf of y . Then, $y_{(1)}, \dots, y_{(n)}$, the ordered observations, are the quantiles $u_{n,j} = (j - 0.5)/n$ of F_n and should be close to $\sigma_0 Q^*(u_{n,j}, \lambda_0) + \mu_0$ for $j = 1, \dots, n$. Consider the differences between the empirical and the theoretical quantiles

$$r_{n,j}(\boldsymbol{\theta}) = y_{(j)} - \mu - \sigma Q^*(u_{n,j}, \lambda), \quad j = 1, \dots, n.$$

The $Q\tau$ estimator is defined by

$$\tilde{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \tau(r_{n,1}(\boldsymbol{\theta}), \dots, r_{n,n}(\boldsymbol{\theta})),$$

where τ denotes the τ scale.

The τ scale was introduced by Yohai and Zamar (1988) to define estimators which combine high finite sample breakdown point with high efficiency in the linear model with normal errors. Given a sample $\mathbf{u} = (u_1, \dots, u_n)$, a function $s(\mathbf{u})$ is called a *scale* if: (i) $s(\mathbf{u}) \geq 0$; (ii) for any scalar γ , $s(\gamma\mathbf{u}) = |\gamma|s(\mathbf{u})$; (iii) $s(u_1, \dots, u_n) = s(|u_1|, \dots, |u_n|)$; (iv) if $|u_i| \leq |v_i|$, $1 \leq i \leq n$, then $s(u_1, \dots, u_n) \leq s(v_1, \dots, v_n)$. It follows that (v) $s(0, \dots, 0) = 0$ and that, (vi) given $\varepsilon > 0$, there exists δ such that $|u_i| \leq \delta$ for $1 \leq i \leq n$ imply $s(u_1, \dots, u_n) < \varepsilon$. Properties (i)-(vi) clearly show that $s(\mathbf{u})$ can be used as a measure of the absolute largeness of the elements of \mathbf{u} . The most common scale is the one based on the quadratic function and is given by $s_1(\mathbf{u}) = (\sum_{i=1}^n u_i^2/n)^{1/2}$. This scale is clearly non robust. Huber (1981) defines a general class of robust scales, called M scales, as follows. Let ρ be a function satisfying the following properties:

- A1** : (i) $\rho(0) = 0$; (ii) ρ is even; (iii) if $|x_1| < |x_2|$, then $\rho(x_1) \leq \rho(x_2)$; (iv) ρ is bounded; (v) ρ is continuous.

Then, an M scale $s_2(\mathbf{u})$ based on ρ is defined by the value s satisfying

$$\frac{1}{n} \sum_{i=1}^n \rho\left(\frac{u_i}{s}\right) = b, \quad (15)$$

where b is a given scalar and $0 < b < a = \sup \rho$. Yohai and Zamar (1988) introduce the family of τ scales. A τ scale is based on two functions ρ_1 and ρ_2 satisfying conditions **A1** and such that $\rho_2 \leq \rho_1$. One considers an M scale $s_2(\mathbf{u})$ defined by (15) with ρ_1 in place of ρ ; then, the τ scale is given by

$$\tau^2(\mathbf{u}) = s_2^2(\mathbf{u}) \frac{1}{n} \sum_{i=1}^n \rho_2 \left(\frac{u_i}{s_2(\mathbf{u})} \right). \quad (16)$$

Usually, ρ , ρ_1 and ρ_2 are selected in the Tukey's bisquare family given by

$$\rho_c^T(u) = 1 - \max \left(\left(1 - \left(\frac{u}{c} \right)^2 \right)^3, 1 \right) \quad (17)$$

for convenient values of c and b (see 5).

The $Q\tau$ estimator in the case of randomly censored observations is obtained by replacing the quantiles of the empirical distribution by the quantiles of the Kaplan-Meier (KM) distribution corresponding to the non censored observations. More precisely, let \tilde{F}_n denote the KM estimator (Kaplan and Meier, 1958) of F_{θ_0} and $t_{(1)}, \dots, t_{(m)}$ the ordered non censored observations. Then, $t_{(1)}, \dots, t_{(m)}$ are the quantiles $\tilde{u}_{n,i} = \tilde{F}_n(t_{(i)}) - 0.5/n$. The residuals

$$\tilde{r}_{n,i}(\boldsymbol{\theta}) = t_{(i)} - \mu - \sigma Q^*(\tilde{u}_{n,i}, \lambda), \quad i = 1, \dots, m.$$

are then used to define the $Q\tau$ estimator for censored observations by

$$\tilde{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \tau(\tilde{r}_{n,1}(\boldsymbol{\theta}), \dots, \tilde{r}_{n,m}(\boldsymbol{\theta})). \quad (18)$$

It is known that the KM estimator distributes the mass of the censored observations among all the observations that are on their right. Some of these observations may be outliers and, therefore, the mass assigned to the outliers by KM may be inflated. To reduce the influence of outliers we defined a trimmed version of the $Q\tau$ estimator as follows. Let $0 < \alpha < 1$ be the fraction of trimming and let $k_\alpha = \max\{h : \tilde{F}_n(t_{(h)}) \leq 1 - \alpha\}$ then the α -trimmed $Q\tau$ (α -TQ τ) estimator is defined by

$$\tilde{\boldsymbol{\theta}}_n = \arg \min_{\boldsymbol{\theta}} \tau(\tilde{r}_{n,1}(\boldsymbol{\theta}), \dots, \tilde{r}_{n,k_\alpha}(\boldsymbol{\theta})). \quad (19)$$

Since in all the simulations and examples we use $\alpha = 0.1$, to simplify the notation in the remaining of the paper, we will write TQ τ instead of α -TQ τ .

Note that the residuals $\tilde{r}_{n,i}(\boldsymbol{\theta})$ are heteroskedastic and, according to Serfling (1980), their variance can be approximated by

$$\hat{\sigma}_i^2 = v_{\tilde{u}_{n,i}}^{*2} / f_{\tilde{\lambda}}^2(Q^*(\tilde{u}_{n,i}, \tilde{\lambda})), \quad (20)$$

where $v_{\tilde{u}_{n,i}}^{*2}$ is Greenwood's variance estimator of $\tilde{F}_n(t_{(i)})$ (Greenwood, 1926). Then, as in Agostinelli et al. (2014a), we might consider the *weighted* TQ τ (WTQ τ) estimator

$$\tilde{\boldsymbol{\theta}}_n^w = \arg \min_{\boldsymbol{\theta}} \tau \left(\frac{\tilde{r}_{n,1}(\boldsymbol{\theta})}{\hat{\sigma}_1}, \dots, \frac{\tilde{r}_{n,k_\alpha}(\boldsymbol{\theta})}{\hat{\sigma}_{k_\alpha}} \right). \quad (21)$$

However, our Monte Carlo experiments have shown that weighting does not provide an important improvement and, for this reason, we are not going to consider the WTQ τ estimator further.

In Theorem 1 of the Appendix we prove that, under general conditions, the TQ τ estimator $\tilde{\boldsymbol{\theta}}_n$ is $n^{1/2}$ -consistent, that is,

$$n^{1/2}(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}_0) = O_p(1). \quad (22)$$

As in the non censored case, one drawback of the quantile τ estimators is that they are not asymptotically normally distributed, making inference difficult. In order to overcome this problem we introduce, in the next Section, the one step truncated ML estimator starting at TQ τ . This estimator has similar robustness properties as the TQ τ estimator; in addition, it has asymptotic normal distribution with the same asymptotic variance as the ML estimator under the model.

3 The truncated maximum likelihood estimators

In order to obtain a robust and highly efficient procedure for estimating the unknown parameter vector we use the truncated maximum likelihood (TML) estimator and the one-step TML (1TML) estimator. Both are based on a weighted form of the likelihood equations. Suppose that $y \sim GLG(\boldsymbol{\theta}_0)$ and consider a sample $(\mathbf{z}_1, \dots, \mathbf{z}_n)$, $\mathbf{z}_i = (y_i^*, \delta_i)$ ($i = 1, \dots, n$). Assume that $\tilde{\boldsymbol{\theta}} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})$ is an initial consistent estimator of $\boldsymbol{\theta}_0$. A natural robustification

of the likelihood equations (14) can be obtained by weighting the equations. More precisely, given a weight function $w(y, \boldsymbol{\theta})$, we consider the equations

$$E_{n, \boldsymbol{\theta}}[w(y, \tilde{\boldsymbol{\theta}}) \mathbf{d}(y, \boldsymbol{\theta})] = E_{n, \tilde{\boldsymbol{\theta}}}[w(y, \tilde{\boldsymbol{\theta}}) \mathbf{d}(y, \tilde{\boldsymbol{\theta}})]. \quad (23)$$

The right hand side mitigates the bias of the estimator and allows to proof asymptotic normality. For increasing sample size its tends to zero. In the next subsections we show how one can define the weight function.

3.1 The outlier rejection rule

We proceed as in Marazzi and Yohai (2004). Let $\tilde{r}_i^* = (y_i^* - \tilde{\mu})/\tilde{\sigma}$ denote the standardized residuals with respect to the initial model and let $l_\lambda(z) = -\log f_\lambda(z)$ be the negative log-likelihood function. We consider the negative log-likelihoods of the residuals $l_i^* = l_{\tilde{\lambda}}(\tilde{r}_i^*)$ ($i = 1, \dots, n$). A large l_i^* corresponds to an observation with a small likelihood under the model and suggests that y_i is an outlier. Let M_λ be the cdf of $l_\lambda(z)$ and

$$M_{n, \tilde{\lambda}}(z) = \frac{1}{n} \sum_{i=1}^n [\delta_i I(l_i^* \leq z) + (1 - \delta_i) P_{\tilde{\lambda}}(l_{\tilde{\lambda}} \leq z | y > y_i^*)],$$

be the semiempirical cdf of l_1^*, \dots, l_n^* for $\lambda = \tilde{\lambda}$. One can show that $M_{n, \tilde{\lambda}}$ is a consistent estimator of M_{λ_0} . For simplicity, we write M_n in place of $M_{n, \tilde{\lambda}}$. Let $M_n^{(\varphi)}$ denote M_n truncated at φ , i.e.,

$$M_n^{(\varphi)}(t) = \begin{cases} M_n(t)/M_n(\varphi) & \text{if } t \leq \varphi, \\ 1 & \text{otherwise.} \end{cases} \quad (24)$$

We want to compare the right tail of the truncated empirical distribution $M_n^{(\varphi)}$ with the right tail of $M_{n, \tilde{\lambda}}(t)$, which is the theoretical distribution when $\lambda = \tilde{\lambda}$. To specify what we understand by the tail, we take a number ε close to 0, for example $\varepsilon = 0.01$, as the probability of falling in the tail. Then, we define the cutoff point φ^* on the likelihood scale as the largest φ such that $M_n^{(\varphi)}(t) \geq M_{\tilde{\lambda}}(t)$ for all $t \geq M_{\tilde{\lambda}}^{-1}(1 - \varepsilon)$, i.e.,

$$\varphi^* = \sup\{\varphi | M_n^{(\varphi)}(t) \geq M_{\tilde{\lambda}}(t) \text{ for all } t \geq M_{\tilde{\lambda}}^{-1}(1 - \varepsilon)\}.$$

Note that φ^* is the minimum value φ such that the tails of $M_n^{(\varphi)}(t)$ and $M_{\tilde{\lambda}}(t)$ are comparable. As in Gervini and Yohai (2002), one can prove that,

if the sample does not contain outliers, $\varphi^* \rightarrow \infty$ a.s. as $n \rightarrow \infty$. Finally, since $l_{\tilde{\lambda}}(z)$ is unimodal, there exist two solutions \tilde{c}_L and \tilde{c}_U of the equation $l_{\tilde{\lambda}}(z) = \varphi^*$. It is immediate that $l_{\tilde{\lambda}}(z) \leq \varphi^*$ is equivalent to $\tilde{c}_L \leq z \leq \tilde{c}_U$. The cutoff points on the data scale are $\tilde{t}_L = \tilde{\mu} + \tilde{\sigma}\tilde{c}_L$ and $\tilde{t}_U = \tilde{\mu} + \tilde{\sigma}\tilde{c}_U$.

3.2 Weight functions

Let $\omega(z)$ be a function, such that

A2 (i) $\omega(z)$ is nonincreasing; (ii) $\lim_{z \rightarrow -\infty} \omega(z) = 1$; (iii) $\omega(z) = 0$ for $z > 0$.

For example, let $c > 0$ and consider the function

$$\omega(z) = \rho(z, c) \cdot I(z \leq 0), \quad (25)$$

where $\rho(z, c) = \rho^T(z/c)$ is in the biweight family as defined in (17) and let $\vartheta^* = 1/\varphi^*$. Then, define the weight function

$$w(z, \lambda, \vartheta^*) = \omega(l_{\lambda}(z) - 1/\vartheta^*)$$

or, for the observation y ,

$$w(y, \boldsymbol{\theta}, \vartheta^*) = w\left(\frac{y - \mu}{\sigma}, \lambda, \vartheta^*\right).$$

Note that for $c \rightarrow 0$, we have $\omega(z) = I(z \leq 0)$. In this case, $w_i = 1$ if $y_i \in [\tilde{t}_L, \tilde{t}_U]$ and $w_i = 0$ otherwise; this rule is usually called *hard rejection*.

3.3 The TML estimators

We suppose that an initial estimator $\tilde{\boldsymbol{\theta}} = (\tilde{\mu}, \tilde{\sigma}, \tilde{\lambda})$ and a cutoff point ϑ^* are given. The *TML estimator* $\hat{\boldsymbol{\theta}} = (\hat{\mu}, \hat{\sigma}, \hat{\lambda})$ is the solution of the equations

$$\mathbf{u}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*) = \tilde{\mathbf{h}}, \quad (26)$$

where $\mathbf{u} = (u_1, u_2, u_3)^\top$ is given by

$$\begin{aligned} \mathbf{u}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*) &= E_{n, \tilde{\boldsymbol{\theta}}} \left[w(y, \tilde{\boldsymbol{\theta}}, \vartheta^*) \mathbf{d}(y, \boldsymbol{\theta}) \right] \\ &= \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\theta}} \left[w(y, \tilde{\boldsymbol{\theta}}, \vartheta^*) \mathbf{d}(y, \boldsymbol{\theta}) | y_i^*, \delta_i \right] \end{aligned}$$

and

$$\tilde{\mathbf{h}} = (\tilde{h}_1, \tilde{h}_2, \tilde{h}_3)^\top = E_{\tilde{\lambda}} \left[w(u, \tilde{\lambda}, \vartheta^*) \mathbf{d}(u) \right] \quad (27)$$

with $u \sim f_{(0,1,\tilde{\lambda})}$. These equations are of the form (23). Note that, when $\vartheta^* \rightarrow 0$, the right hand side vector tends to zero and we obtain the ML equations.

The *1TML estimator* is obtained by applying one iteration of the Newton-Raphson procedure to equations (26) and it turns out to be

$$\hat{\boldsymbol{\theta}}_1 = \tilde{\boldsymbol{\theta}} - \mathbf{J}(\tilde{\boldsymbol{\theta}}, \vartheta^*)^{-1} (\mathbf{u}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*) - \tilde{\mathbf{h}}), \quad (28)$$

where $\mathbf{J}(\tilde{\boldsymbol{\theta}}, \vartheta^*) = \nabla_{\boldsymbol{\theta}} \mathbf{u}(\boldsymbol{\theta}, \tilde{\boldsymbol{\theta}}, \vartheta^*)|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$ is a Jacobian matrix. Similarly, we can define a *two-step TML* (2TML) *estimator* $\hat{\boldsymbol{\theta}}_2$ by replacing $\tilde{\boldsymbol{\theta}}$ with $\hat{\boldsymbol{\theta}}_1$, $\tilde{\mathbf{h}}$ with $\hat{\mathbf{h}}_1$, and $\mathbf{J}(\tilde{\boldsymbol{\theta}})$ with $\mathbf{J}(\hat{\boldsymbol{\theta}}_1)$ in (28).

In Theorem 4 of the Appendix we prove that, under general conditions including $n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) = O_p(1)$, the 1TML estimator satisfies the following asymptotic result:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{G}(\boldsymbol{\theta}_0)^{-1} \mathbf{M}(\boldsymbol{\theta}_0) \mathbf{G}(\boldsymbol{\theta}_0)^{-\top}).$$

A similar result holds for the 2TML estimator. In the Monte Carlo simulations, reported in Section 5, we show that for finite sample sizes, the 1TML and the 2TML estimators are more efficient than the TQ τ estimator (with 10% trimming) and have a reasonably robust behavior under outlier contamination. Therefore we propose, as a final procedure to estimate $\boldsymbol{\theta}_0$, the 2TML estimator starting with the TQ τ estimator with 10% trimming. Numerical experiments show that further steps do not provide any significant improvement.

4 The regression case

We now consider an AFT model for pairs of observations (\mathbf{x}_i, y_i) , $1 \leq i \leq n$, where $\mathbf{x}_i \in \mathbb{R}^p$ and $y_i \in \mathbb{R}$ satisfying

$$y_i = \mu_0 + \boldsymbol{\beta}_0^\top \mathbf{x}_i + \sigma_0 u_i, \quad i = 1, \dots, n, \quad (29)$$

where y_i may represents a duration on the logarithmic scale and \mathbf{x}_i the corresponding covariables vector. The slopes $\boldsymbol{\beta}_0 \in \mathbb{R}^p$, the intercept μ_0 , and the scale σ_0 are unknown parameters. The errors u_i , $1 \leq i \leq n$, are assumed to

be i.i.d. and independent of \mathbf{x}_i . Moreover, the distribution of the carriers \mathbf{x}_i is unknown. We assume that the error density is $f_{(0,1,\lambda_0)}$, where λ_0 is an unknown shape parameter and $f_{\boldsymbol{\theta}}$ is given by (3). We observe $(y_i^*, \mathbf{x}_i, \delta_i)$, where $y_i^* = \min(y_i, c_i)$, c_1, \dots, c_n are i.i.d. censoring times, which are independent of the u_i 's. We put $\delta_i = 1$ if $y_i^* = y_i$ and $\delta_i = 0$ otherwise. We write $\boldsymbol{\gamma}_0 = (\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)$, where $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$.

Let $\boldsymbol{\gamma} = (\boldsymbol{\theta}, \boldsymbol{\beta})$ and $u = (y - \mu - \boldsymbol{\beta}^\top \mathbf{x})/\sigma$. Let $\mathbf{d}(\mathbf{x}, y, \boldsymbol{\gamma})$ and $\mathbf{s}(\mathbf{x}, y, \boldsymbol{\gamma})$ denote the $(3+p)$ -component non-censored and censored regression score function vectors respectively. The first 3 components of $\mathbf{d}(\mathbf{x}, y, \boldsymbol{\gamma})$ and $\mathbf{s}(\mathbf{x}, y, \boldsymbol{\gamma})$ are given by the right-hand sides of (5)-(7) and (8)-(10) respectively. In addition, for $k = 1, \dots, p$, we have

$$\begin{aligned} d_{3+k}(\mathbf{x}, y, \boldsymbol{\gamma}) &= \frac{x_k}{\sigma} \xi_\lambda(u), \\ s_{3+k}(\mathbf{x}, y, \boldsymbol{\gamma}) &= -\frac{x_k}{\sigma} \frac{f_\lambda(u)}{S_\lambda(u)}. \end{aligned}$$

Then, simple derivations shows that the ML estimator of $\boldsymbol{\gamma}$ is the solution of

$$\frac{1}{n} \sum_{j=1}^n \mathbf{v}(\mathbf{x}_j, y_j^*, \delta_j, \boldsymbol{\gamma}) = \mathbf{0}, \quad (30)$$

where $\mathbf{v}(\mathbf{x}, y, \delta, \boldsymbol{\gamma}) = \delta \mathbf{d}(\mathbf{x}, y, \boldsymbol{\gamma}) + (1 - \delta) \mathbf{s}(\mathbf{x}, y, \boldsymbol{\gamma})$. A similar expression as (14) can be obtained, where the *semiparametric cdf* is defined by

$$\begin{aligned} H_{n,\boldsymbol{\gamma}}(z, \mathbf{z}) &= \frac{1}{n} \sum_{i=1}^n E_{\boldsymbol{\gamma}} [I(y \leq z) | y_i^*, \mathbf{x}_i, \delta_i] I(\mathbf{x}_i \leq \mathbf{z}) \\ &= \frac{1}{n} \sum \delta_i I(y_i \leq z) I(\mathbf{x}_i \leq \mathbf{z}) \\ &\quad + \frac{1}{n} \sum (1 - \delta_i) \frac{[F_{\boldsymbol{\gamma}}(z) - F_{\boldsymbol{\gamma}}(y_i^*)]^+}{1 - F_{\boldsymbol{\gamma}}(y_i^*)} I(\mathbf{x}_i \leq \mathbf{z}). \end{aligned}$$

We denote by $E_{n,\boldsymbol{\gamma}}$ the expectation with respect to $H_{n,\boldsymbol{\gamma}}$. In the next two subsections we define a robust and efficient procedure for estimating $\boldsymbol{\gamma}$. In a first step an initial highly robust but not necessarily efficient estimator of $\boldsymbol{\gamma}_0$ is computed; the second step uses a highly robust and asymptotically efficient estimator of $\boldsymbol{\gamma}_0$ based on a 1TML procedure.

4.1 The initial regression estimator

An initial estimator of γ_0 can be defined as follows:

1. Let $\bar{\mu}$ and $\bar{\beta}$ be MM-estimators for censored data of μ_0 and β_0 as defined in Salibian-Barrera and Yohai (2008).
2. Let $\omega_j = y_j - \bar{\beta}^\top \mathbf{x}_j$ and $\omega_j^* = y_j^* - \bar{\beta}^\top \mathbf{x}_j = \min(\omega_j, c_j^*)$ where $c_j^* = c_j - \bar{\beta}^\top \mathbf{x}_j$. For large n , the distribution of the ω_j s is close to F_{θ_0} for $1 \leq j \leq n$. We can then estimate θ_0 using the observations $(\omega_1^*, \delta_1), \dots, (\omega_n^*, \delta_n)$ by means of a TQ τ estimator that will be denoted by $\tilde{\theta} = (\tilde{\mu}, \tilde{\sigma}, \lambda)$.
3. The initial regression estimator of γ_0 is $\tilde{\gamma} = (\tilde{\theta}, \bar{\beta})$. It will be called *MM-TQ τ estimator*.

In Theorem 2 we show that, under general conditions, if $\bar{\beta}$ is $n^{1/2}$ consistent, that is if

$$n^{1/2}(\bar{\beta} - \beta_0) = O_p(1), \quad (31)$$

then $\tilde{\theta}$ satisfies

$$n^{1/2}(\tilde{\theta} - \theta_0) = O_p(1). \quad (32)$$

The result (31) remains still a conjecture, however Salibian-Barrera and Yohai (2008) provide compelling arguments in favor of this conjecture (see in particular their Theorem 6). Besides, their Monte Carlo study seems to confirm this conjecture.

4.2 The final regression estimator

Let $\mu(\mathbf{x}) = \mu + \mathbf{x}^\top \beta$ and $\tilde{\mu}(\mathbf{x}) = \tilde{\mu} + \mathbf{x}^\top \bar{\beta}$. Then, the standardized residuals with respect to the initial estimator are $\tilde{r}_i^* = (y_i^* - \tilde{\mu}(\mathbf{x}_i))/\tilde{\sigma}$ and can be used to obtain the cutoff point ϑ^* and the weights

$$w(\mathbf{x}, y, \gamma, \vartheta) = w\left(\frac{y - \tilde{\mu}(\mathbf{x})}{\tilde{\sigma}}, \tilde{\lambda}, \vartheta\right).$$

The *TML regression estimator* $\hat{\gamma}$ is the solution of the equations

$$\mathbf{u}(\gamma, \tilde{\gamma}, \vartheta^*) = \tilde{\mathbf{h}}, \quad (33)$$

where

$$\mathbf{u}(\gamma, \tilde{\gamma}, \vartheta^*) = E_{n, \tilde{\gamma}}[w(\mathbf{x}, y, \tilde{\gamma}, \vartheta^*)\mathbf{d}(\mathbf{x}, y, \gamma)],$$

and $\tilde{\mathbf{h}} = (\tilde{h}_1, \tilde{h}_2, \tilde{h}_3, \tilde{\mathbf{h}}_4^\top)^\top$, where $\tilde{h}_1, \tilde{h}_2, \tilde{h}_3$ are defined in (27) and

$$\tilde{\mathbf{h}}_4 = \frac{1}{n} \sum_{i=1}^n E_{\tilde{\lambda}} \left[w(u, \tilde{\lambda}, \vartheta^*) \xi_{\tilde{\lambda}}(u) \mathbf{x}_i \right].$$

The *1TML regression estimator* is obtained by applying one Newton-Raphson iteration to equations (33). This estimator turns out to be

$$\hat{\gamma}_1 = \tilde{\gamma} - \mathbf{J}(\tilde{\gamma}, \vartheta^*)^{-1}(\mathbf{u}(\gamma, \tilde{\gamma}, \vartheta^*) - \tilde{\mathbf{h}}),$$

where $\mathbf{J}(\tilde{\gamma}, \vartheta^*) = \nabla_{\gamma} \mathbf{u}(\gamma, \tilde{\gamma}, \vartheta^*)|_{\gamma=\tilde{\gamma}}$ is a Jacobian matrix. The *2TML regression estimator* is defined in the obvious way.

In Theorem 3 of the Appendix we prove that, under general conditions including $n^{1/2}(\tilde{\gamma} - \gamma_0) = O_p(1)$, the 1TML regression estimator satisfies the following asymptotic result:

$$\sqrt{n}(\hat{\gamma}_1 - \gamma_0) \xrightarrow{L} N(\mathbf{0}, \mathbf{G}(\gamma_0)^{-1} \mathbf{M}(\gamma_0) \mathbf{G}(\gamma_0)^{-\top}),$$

where $\mathbf{M}(\gamma) = E(\mathbf{v}(\mathbf{x}, y, \delta, \gamma) \mathbf{v}(\mathbf{x}, y, \delta, \gamma)^\top)$ and $\mathbf{G}(\gamma) = E(\nabla_{\gamma} \mathbf{v}(\mathbf{x}, y, \delta, \gamma))$. The same result can be obtained for the 2TML regression estimator.

5 Monte Carlo experiments

5.1 The case without covariates

A first set of experiments was run in the case without covariates. In general, the results are similar to those described in Agostinelli et al. (2014a) for the non censored case and we report here only the most representative cases. We compared the following estimators: ML, TQ τ , 1TML, 2TML, and the one step WL estimator (1SWL) as defined in Section A.3.1 of the Appendix.

The TQ τ estimator was defined using ρ_1 and ρ_2 in the Tukey's bisquare family (17) with c equal to 1.548 and 6.08 respectively, and $b = 0.5$. The values 1.548 and $b = 0.5$ have been chosen so that the regression estimator based on the τ -scale has a finite sample breakdown point equal to 0.5. The value 6.08 makes the asymptotic efficiency equal to 0.95 in the case of normal errors. To compute the 1SWL estimator, we used a normal kernel with bandwidth $h = 0.3\hat{\sigma}$ in all experiments, where the scale estimator $\hat{\sigma}$ is the scale initial estimate.

To compare the global performances of the different estimators we consider, for each set of parameters $\boldsymbol{\theta} = (\mu, \sigma, \lambda)$, the total variation distance (TVD)

$$\text{TVD}(\boldsymbol{\theta}) = \int |f_{\boldsymbol{\theta}}(y) - f_{\boldsymbol{\theta}_0}(y)| dy$$

between the density $f_{\boldsymbol{\theta}}$ and the true underlying density $f_{\boldsymbol{\theta}_0}$. The performance of the estimator $\hat{\boldsymbol{\theta}}$ is measured by the mean value of $\text{TVD}(\hat{\boldsymbol{\theta}})$, that is by

$$\text{MTVD}(\hat{\boldsymbol{\theta}}) = E(\text{TVD}(\hat{\boldsymbol{\theta}})).$$

$\text{MTVD}(\hat{\boldsymbol{\theta}})$ clearly measures the quality of the estimated density. It is estimated, using the simulated values $\hat{\boldsymbol{\theta}}_k$ ($1 \leq k \leq N$) of the estimator, by

$$\text{ATVD}(\hat{\boldsymbol{\theta}}) = \frac{1}{N} \sum_{k=1}^N \text{TVD}(\hat{\boldsymbol{\theta}}_k).$$

5.1.1 Simulation under the nominal model

We studied the efficiency of the estimators under the nominal model for $n = 50, 100, 400$ and 1000 and for λ_0 equal 1 and 2 . Without loss of generality we took $\mu_0 = 0$ and $\sigma_0 = 1$. We considered two censoring fractions: 15% and 25% . The number of replications was 1200 . Figure 1 (top) reports the ATVD of the robust estimators divided by the ATVD of the ML estimator as a function of the sample size for $\lambda = 1$ and censoring fraction of 15% . This ratio can be interpreted as a measure of relative efficiency that we call *TVD efficiency*. As expected, the TVD efficiency of 1TML, 2TML and 1SWML is markedly larger than the efficiency of the initial estimator as the sample size grows. Moreover, when n increases the TVD efficiency becomes close to 1 , that is, it becomes close to the efficiency of the ML estimator. Similar patterns are observed in the other simulated cases.

5.1.2 Simulation under point mass contamination

In a second Monte Carlo experiment, we compared TQ τ with 10% trimming, 1TML, 2TML and 1SWL under point mass contamination for $n = 50, 100, 400, 1000$, $\lambda_0 = 1$, $\lambda_0 = 2$ and two censoring fractions: 15% and 25% . We generated 90% “good” observations y_j according to the GLG model and 10% “outliers” at the point y_0 . We then varied the value of y_0 from -10 to 10

with a step of 0.5. This kind of point mass contamination is generally the least favorable one and allows evaluation of the maximal bias an estimator can incur. For each value of y_0 , the number of replications was 1200. Figure 1 (bottom) reports the average TVD (ATVD) of the estimated densities as a function of y_0 . The results show that the ATVD of TQ τ , 1TML and 2TML are comparable and very stable over the whole range of y_0 values, while 1SWL provides a very high ATVD for positive values of y_0 . The ATVD of ML is not shown because of its unbounded behavior.

5.2 Regression case

A second set of experiments was run to investigate the behavior of the MM-TQ τ estimator defined in Section 4.1, the 1TML and 2TML estimators defined in Section 4.2, and the 1SWL estimator defined in Section A.3.2 of the Appendix.

In each experiment, n pairs of observations (\mathbf{x}_i, y_i^*) were generated according to the regression model

$$y_i = \mu_0 + \beta_0^\top \mathbf{x}_i + \sigma_0 u_i, \quad (34)$$

$$y_i^* = \min(y_i, c_i), \quad (35)$$

$$c_i = \mu_c + e_i, \quad (36)$$

$$\delta_i = 1 \text{ if } y_i^* = y_i \text{ and } 0 \text{ otherwise,} \quad (37)$$

with $\mu_0 = 0$, $\sigma_0 = 1$, $\lambda_0 = 1$ and 2, $\beta_0^\top = (2, 3)$, $\mathbf{x}_i = (x_{i1}, x_{i2})$, where $x_{i1} \sim N(0, 1)$, $x_{i2} \sim \mathcal{B}(1, 0.4)$ with x_{i1} and x_{i2} independent and $e_i \sim GLG(\mu_c, 1, 1)$. The parameter μ_0 was chosen so that the censoring fraction was 15% or 25%.

To compare the estimators we defined, for each set of parameters $(\boldsymbol{\theta}, \boldsymbol{\beta})$ and covariable vector \mathbf{x} ,

$$\text{TVD}(\boldsymbol{\theta}, \boldsymbol{\beta}, \mathbf{x}) = \int |f_{(\boldsymbol{\theta}, \boldsymbol{\beta})}(y|\mathbf{x}) - f_{(\boldsymbol{\theta}_0, \boldsymbol{\beta}_0)}(y|\mathbf{x})| dy,$$

where

$$f_{(\boldsymbol{\theta}, \boldsymbol{\beta})}(y|\mathbf{x}) = f_\lambda((y - \mu - \boldsymbol{\beta}^\top \mathbf{x})/\sigma).$$

Then, we measured the performance of the estimator $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$ by the mean value of $\text{TVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}, \mathbf{x})$, where \mathbf{x} is independent of the sample used to compute $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$, that is by

$$\text{MTVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) = E(\text{TVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}, \mathbf{x})).$$

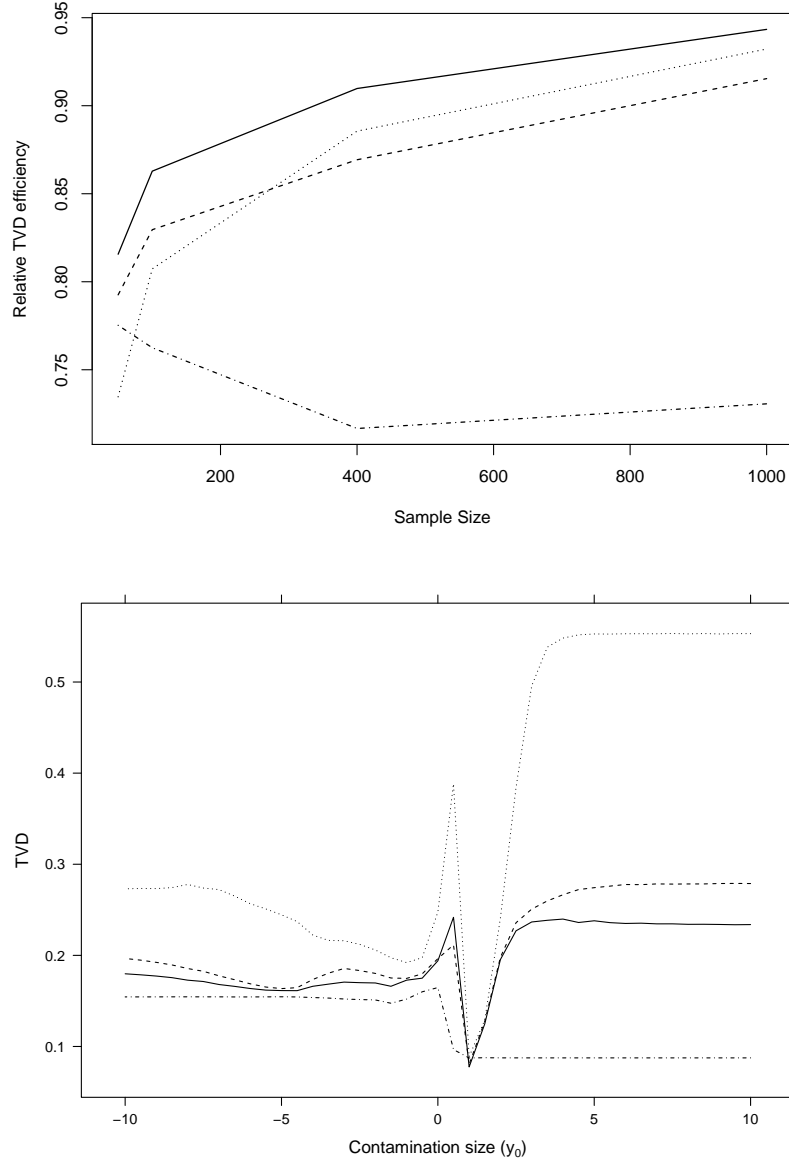


Figure 1: Case without regressors. Top: Estimated relative TVD efficiency versus sample size. Bottom: Average TVD under 10% contamination and sample size $n = 100$. Legend: TQ τ (dashed and dotted line), 1TML (dashed line), 2TML (solid line) and 1SWL (dotted line). The censoring proportion is 15% and $\lambda = 1$.

The expectation was taken on $(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}})$ and \mathbf{x} and estimated using the simulated values $(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\beta}}_k)$ and $\mathbf{x}_{k1}^*, \dots, \mathbf{x}_{ki}^*, \dots, \mathbf{x}_{kn}^*$ ($1 \leq k \leq N$):

$$\text{ATVD}(\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\beta}}) = \frac{1}{Nn} \sum_{k=1}^N \sum_{i=1}^n \text{TVD}(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\beta}}_k, \mathbf{x}_{ki}^*).$$

5.2.1 Simulation under the nominal model

We studied the efficiency of the estimators under the nominal model for $n = 50, 100, 400$ and 1000 . Figure 2 (top) shows the relative TVD efficiency of MM-TQ τ , 1TML, 2TML and 1SWL with respect to ML for $\lambda_0 = 1$, $\mu_0 = 0$, $\sigma_0 = 1$, and censoring fraction 0.15%.

The efficiency of 1TML, 2TML and 1SWL is clearly higher than the efficiency of MM-TQ τ . This is an expected result, since 1TML, 2TML and 1SWL are asymptotically fully efficient under the nominal model.

5.2.2 Simulation under point mass contamination

We also studied the behavior of the estimators under point mass contamination for $n = 50, 100, 400, 1000$. The values of the parameters were the same as in the case of no contamination. We generated n “good” observations (\mathbf{x}_i, y_i^*) according to (34)-(37). We then replaced 10% values y_i^* with a value y_0 ranging from -10 to 20 . For each value of y_0 the number of replications was 1200. Figure 2 (bottom) shows ATVD of the estimators as a function of y_0 for sample size $n = 100$ and contamination level 10%. We observe that MM-TQ τ , 1TML, and 2TML are very resistant under point mass contamination, while the 1SWL is highly sensitive to outliers on the right tail of the distribution.

5.2.3 Empirical finite sample breakdown point

We were not able to obtain the breakdown point of the proposed estimators. To fill this gap, we performed a Monte Carlo simulation to explore the behavior of the maximum mean square error as a function of the contamination level. This provides information about the highest contamination the proposed estimators could cope with and hence about the finite sample breakdown point. We used the same setting as in the previous subsection, $n = 1000$, $\lambda_0 = 1$, and censoring fraction 0.15. Several values of the contamination level ϵ in the interval $[0, 0.3]$ and several values y_0 in the interval

$[-100, 100]$ were considered. For each pair (ϵ, y_0) , we run 100 Monte Carlo replications and computed the maximum MSE (MMSE) of the regression parameters (μ, β) –note that μ is the intercept–, the scale parameter (σ) and the shape parameter (λ). Results for the regression parameters (slopes) are reported in Figure 3; they show that the MMSE starts to increase rapidly around the 20% level. This behavior is consistent with the MMSE of the initial regression parameters provided by the MM non parametric estimator.

6 Illustrations with real data

In modern hospital management, stays are classified into “diagnosis related groups” (DRGs; Fetter et al. (1980) which are designed to be as homogeneous as possible with respect to diagnosis, treatment, and resource consumption. The “mean cost of stay” of each DRGs is periodically estimated with the help of administrative data on a national basis and used to determine “standard prices” for hospital funding and reimbursement. Since it is difficult to measure cost, “length of stay” (LOS) is often used as a proxy. In designing and “refining” the groups, the relationship between LOS and other variables which are usually available on administrative files has to be assessed and taken into account. We discuss two examples in this domain.

6.1 Major cardiovascular interventions

In a first example, we consider a sample of 75 stays in a particular hospital and DRG “Major cardiovascular interventions”. Of these stays, 45 were censored because the patients were transferred to a different hospital before dismissal. The data – shown in Figure 5 and made available in Marazzi and Muralti (2013) – were first analyzed in Locatelli et al. (2010) and Locatelli and Marazzi (2013). These authors studied the relationship between LOS and two covariates: Age of the patient (x_1) and Admission type ($x_2 = 0$ for planned admissions, $x_2 = 1$ for emergency admissions) with the help of the model $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \gamma x_1 x_2 + \sigma u$, where $y = \log(\text{LOS})$ and u is following a Gaussian model. They observed that two young patients had exceptionally high non censored LOS and, as a consequence, the ML estimator yielded an unexpected large estimate of the interaction γ . Therefore, they proposed the use of a robust parametric procedure called “weighted maximum likelihood” (WML/G) based on the Gaussian error model (that

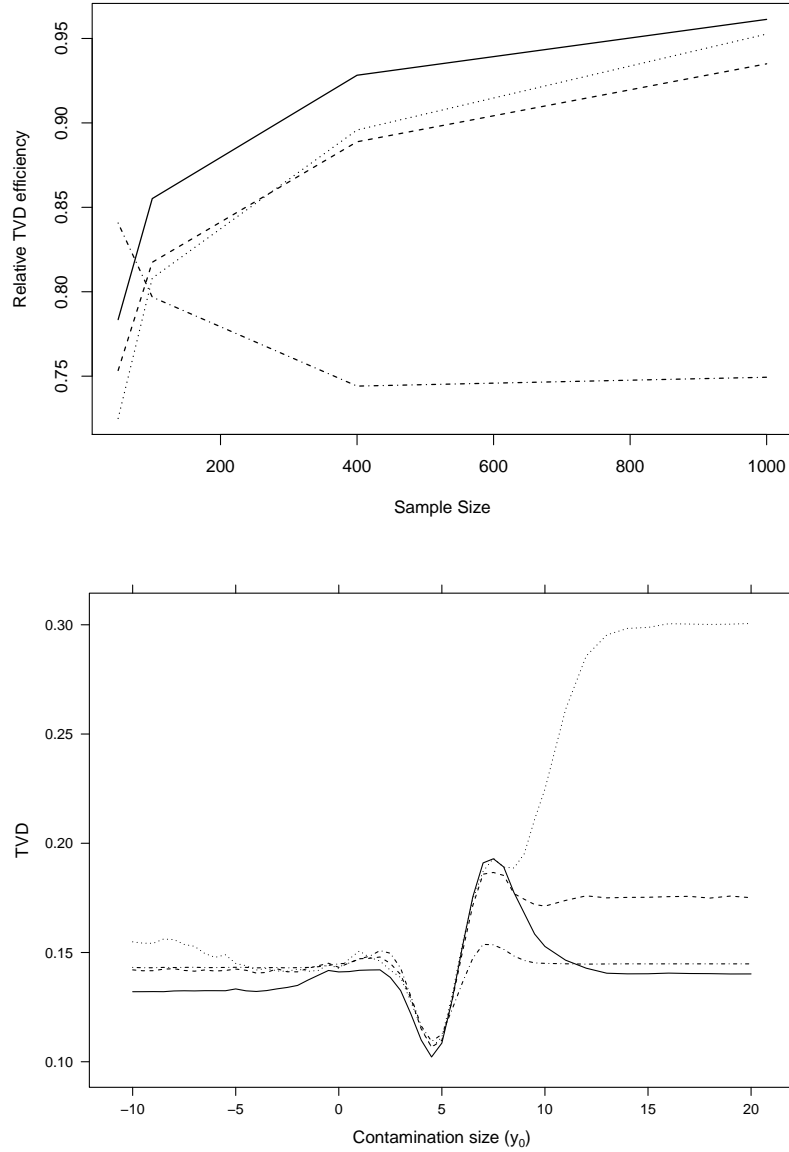


Figure 2: Regression case. Top: Estimated relative TVD efficiency versus sample size. Bottom: Average TVD under 10% contamination and sample size $n = 100$. Legend: MM-TQ τ (dashed and dotted line), 1TML (dashed line), 2TML (solid line) and 1SWL (dotted line). The censoring proportion is 15% and $\lambda = 1$.

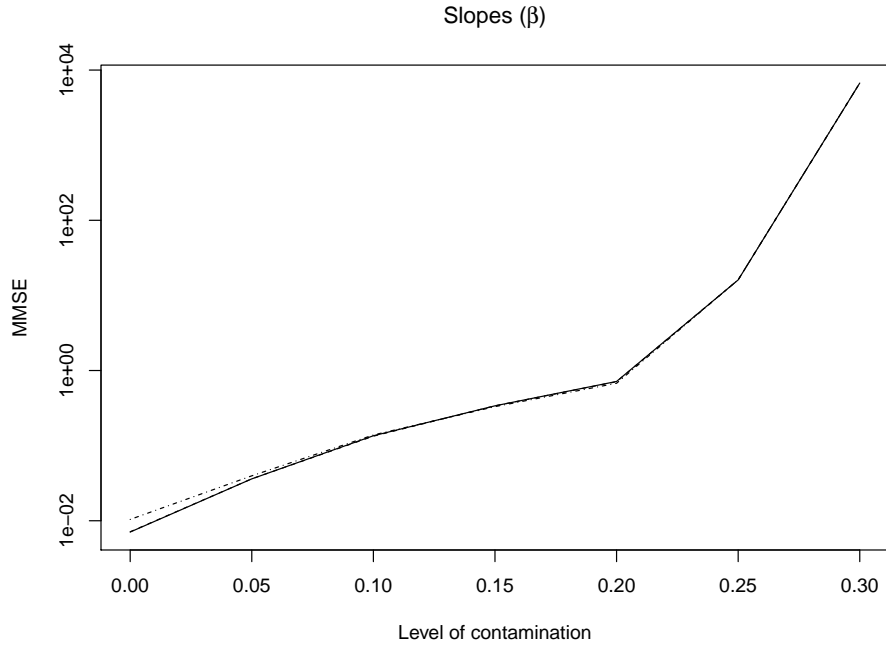


Figure 3: Estimated maximum MSE of the regression parameters (slopes) for MM-TQ_r (dashed and dotted line), 1TML (dashed line), and 2TML (solid line). The censoring proportion is 15%, $n = 1000$ and $\lambda = 1$. The MMSEs of the 1TML and 2TML are almost always overlapping. The y-axis is on log scale.

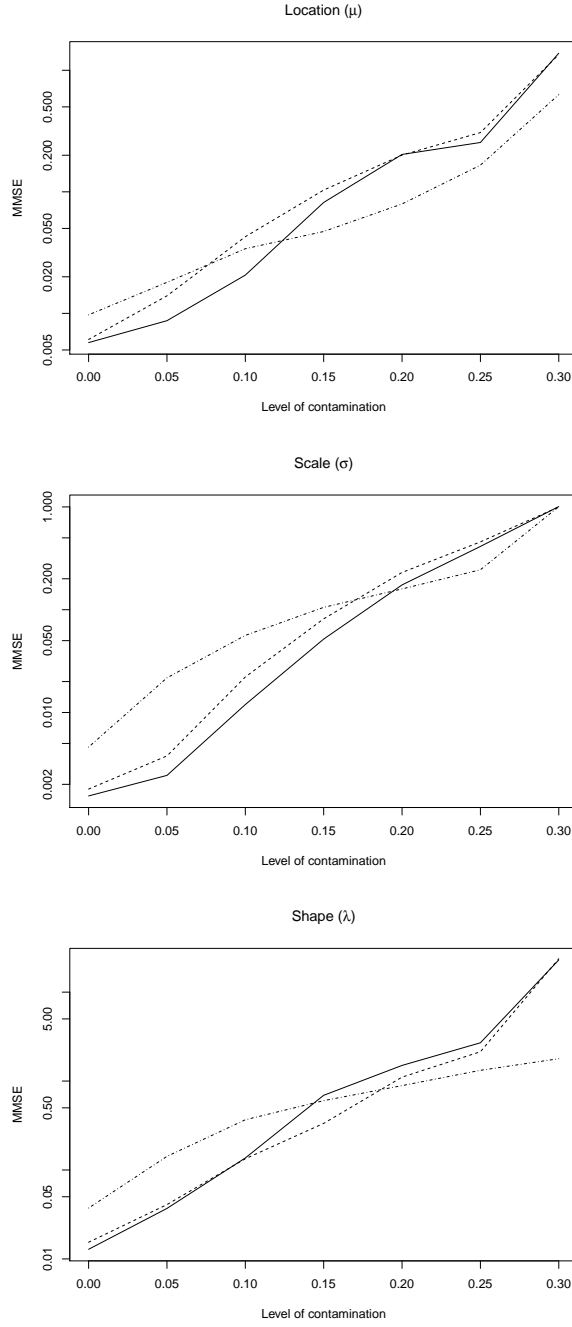


Figure 4: Regression case. Estimated maximum MSE as function of contamination level for TQ τ (dashed and dotted line), 1TML (dashed line) and 2TML (solid line). Top: location parameter, middle: scale parameter, bottom: shape parameter. The censoring proportion is 15%, $n = 1000$ and $\lambda = 1$.

performed better than log-Weibull). They compared WML/G with other published robust procedures and found that the robust estimates of γ were close to zero.

Here, we assume a GLG error model and consider the ML, 1TML, 2TML, 1SWL and 2SWL regression estimates reported in Table 1. Estimated regression lines are reported in Figure 5 as well. For comparison, we also report the ML estimate with Gaussian errors (ML/G) and WML/Gauss. We first notice the good agreement among the robust coefficient estimates based on GLG. The one and two steps TML and WL estimates provide the same inferences as ML after removal of the outliers.

Apart from μ , σ and β_2 , ML yields larger absolute values for the estimates; however, none of them are significant because standard errors are inflated by the outliers. Not surprisingly, the main differences between the robust procedures based on GLG and those based on the Gaussian model concern the intercept terms (μ and β_1); however, the robust prediction lines based on GLG (Figure 5) seem to provide a better fit to the bulk of the data. This observation is supported by the plots in Figure 6, where three types of distributions of the standardized residuals are displayed: KM, parametric (normal and GLG), and semiparametric (normal and GLG). Note the very large steps of KM corresponding to the two extreme non censored observations. The reason is that KM puts the mass of several censored residuals on these two points. The ML survival functions are strongly affected by these two points (see Figure 4a in Locatelli et al. (2010)). Both WML/Gauss and 2TML distribution functions behave much better: with two exceptions, their residuals follow the models very well. However 2TML is clearly better in the left tail. Finally, we note that the use of GLG provides a reasonable fit for the two young patients with high non censored LOS, while the censored observations corresponding to emergency admission in the right bottom corner are considered outliers.

6.2 Minor bladder interventions

In a second example, we consider a sample of 48 stays for DRG “Minor bladder interventions”. The data are shown in Figure 7. Six patients were transferred to a different hospital before dismissal. Four young patients have surprisingly large values of LOS. We study the relationship between LOS and Age considering the model $y = \mu + \beta x + \sigma u$, where $x = \text{Age}$ and $y = \log(\text{LOS})$. The ML, the 1TML, 2TML and 1SWL estimates and estimated standard

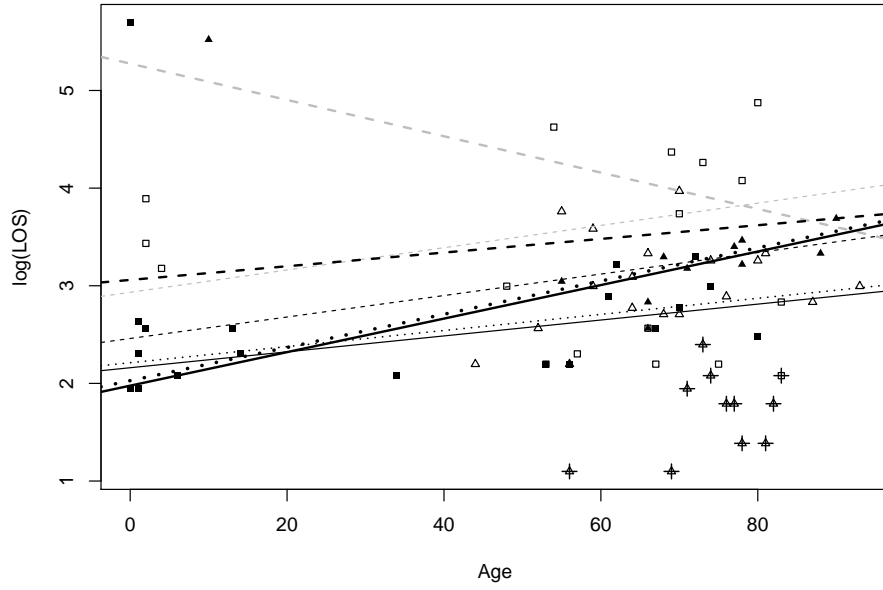


Figure 5: Estimates of the regression models for length of stay of “Major cardiovascular interventions”. Thin lines refer to planned admissions, while thick lines refer to emergency admission. ML is grey and dashed, WML/G is black and dashed, 2TML is black and solid, 2SWL is black and dotted. Filled marks represent complete observation; empty marks represent censored observation; square marks are planned admission; triangle are emergency admission. Outliers according to 2TML are marked with crosses.

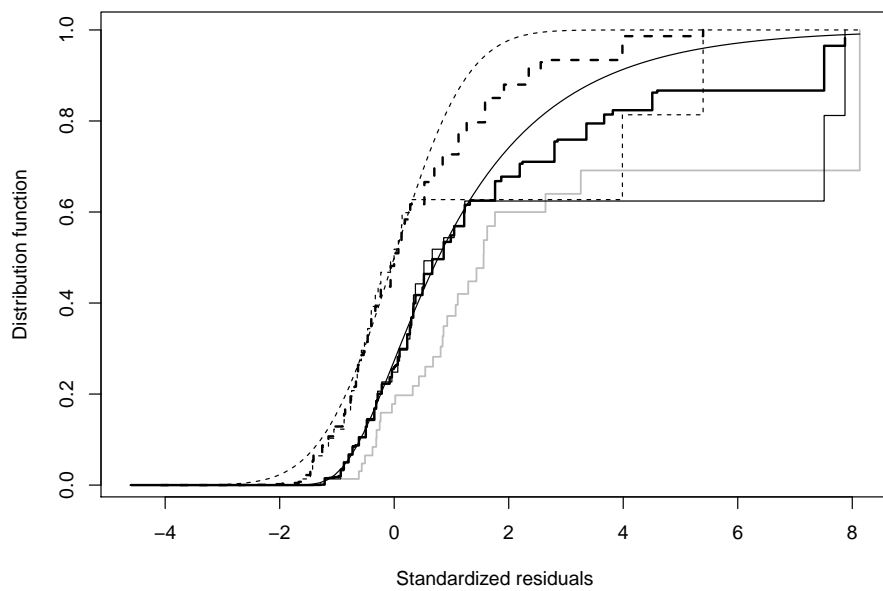


Figure 6: Standardized residuals cdf of different estimates for the “Major cardiovascular interventions” data. Kaplan Meyer cdf of the ML residuals is grey; WML/G residual cdf is black and dashed; 2TML residual cdf is black and solid. Thin lines are Kaplan-Meier cdf; thick lines are semiparametric cdf; the smooth line is the parametric cdf.

Table 1: Estimates of the regression model for length of stay of “Major cardiovascular interventions”. Boldface: p-values smaller than 0.05 for the null hypothesis that the parameter equals zero. Abbreviations: c.d.= complete data set, o.r.= outliers removed, /G= Gaussian errors.

	μ	$10\beta_1$	$10\beta_2$	10γ	σ	λ
1TML	2.16 (0.11)	-1.84 (8.73)	0.08 (0.02)	0.09 (0.12)	0.45 (0.08)	- 1.82 (0.14)
2TML	2.16 (0.11)	-1.84 (8.71)	0.08 (0.02)	0.09 (0.12)	0.45 (0.08)	- 1.82 (0.14)
1SWL	2.16 (0.16)	-1.84 (8.10)	0.08 (0.03)	0.09 (0.12)	0.45 (0.07)	- 1.82 (0.43)
2SWL	2.16 (0.16)	-1.84 (8.12)	0.08 (0.03)	0.09 (0.12)	0.45 (0.07)	- 1.82 (0.43)
ML c.d.	2.05 (1.10)	22.03 (123.00)	0.07 (0.13)	-0.27 (2.04)	0.45 (1.59)	-3.34 (17.54)
ML o.r.	2.09 (0.10)	1.11 (10.02)	0.07 (0.02)	0.04 (0.14)	0.48 (0.10)	- 2.79 (0.58)
ML/G	2.93	23.42	0.11	-0.30	-	-
WML/G	2.44	10.57	0.11	-0.10	-	-

Table 2: Estimates of the regression model for length of stay of “Minor bladder interventions”. Abbreviations: c.d.= complete data set; o.r.= outliers removed.

Method	μ	10β	σ	λ
1TML	0.992 (0.365)	0.127 (0.056)	0.480 (0.050)	0.489 (0.170)
2TML	0.992 (0.365)	0.127 (0.056)	0.482 (0.047)	0.489 (0.172)
1SWL	0.992 (0.600)	0.130 (0.080)	0.481 (0.063)	0.490 (0.120)
2SWL	0.992 (0.606)	0.131 (0.081)	0.484 (0.064)	0.489 (0.120)
ML c.d.	2.633 (0.491)	-0.113 (0.056)	0.686 (0.066)	0.049 (0.565)
ML o.r.	0.525 (0.363)	0.161 (0.052)	0.564 (0.069)	-0.449 (0.339)

errors are reported in Table 2; the corresponding prediction lines are drawn in Figure 7.

According to ML, $\log(\text{LOS})$ does not seem to depend on Age (the p-value is 0.0444) and a Gauss distribution seems to be adequate (the p-value for λ is 0.931). The robust estimates are similar and provide a much larger slope (p-value = 0.019). Moreover – as it is expected from this data – they suggest a positive linear relationship and an asymmetric error model (the p-value for λ is 0.0024). Clearly, the outliers (those with weights equals to zero in 2TML are marked with crosses in Figure 7) have an important leverage effect on the ML coefficients and shape parameter. Removing the outliers, ML becomes similar to the robust estimates. In practice, this simple analysis suggests that the possibility of splitting this particular DRG into two groups should

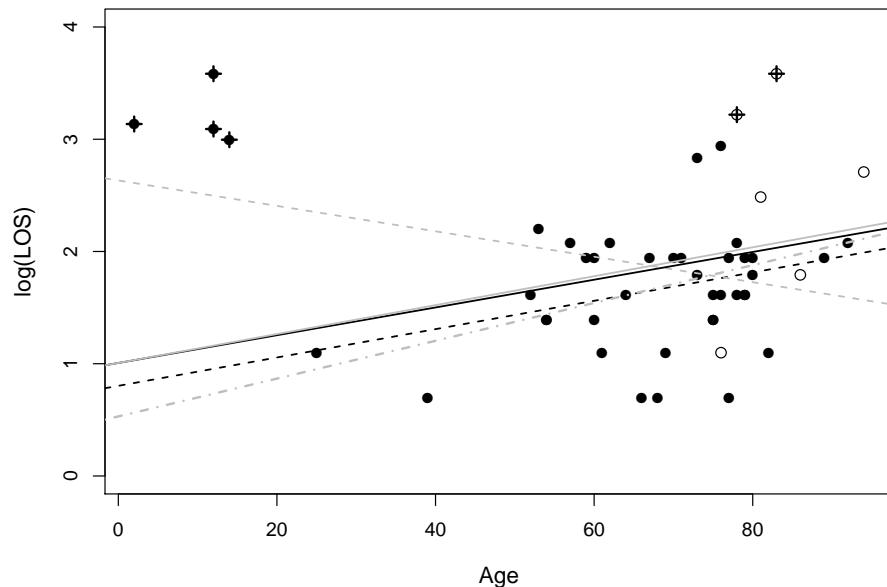


Figure 7: Estimates of the regression models for length of stay of “Major bladder interventions”. Maximum Likelihood based on the complete data set (ML c.d.) is grey and dashed, initial non parametric MM is black and dashed, Maximum Likelihood after removing outliers according to 2TML (ML o.r.) is gray and dash and dotted, 2TML is black and solid and 2SWL is gray and solid. Outliers according to 2TML are marked with a cross.

be further investigated.

7 Concluding remarks

As mentioned in the introduction, the GLG model is a very flexible family of distributions which is used to describe asymmetrically distributed data in many real applications. In this paper, we considered estimators which are simultaneously robust and efficient for AFT models, when the errors follow a GLG distribution and the data may contain censored observations. The estimation procedures have two main components: an initial highly robust

but not necessarily efficient estimator and a final efficient estimator which starts with the initial one.

We first considered the case, where no covariables are present and, in this case, we proposed an initial estimator that minimizes a τ scale of the differences between theoretical and empirical quantiles of order smaller than $(1 - \alpha)$, where $0 < \alpha < 1$ is a trimming fraction. The final estimator is a one step weighted likelihood estimator, where the weights penalizing the outliers are derived from the initial estimator.

For the case, where covariables are present, the proposed estimators were derived in three steps. In a first step we used a regression MM-estimator as proposed in Salibián-Barrera and Yohai (2008) to obtain initial slope estimates and to compute the corresponding residuals. In a second step, we computed an initial estimator of the GLG parameters by applying the procedure for the no covariables case to these residuals. In the third step we obtained a final estimator of all the parameters using a one step truncated ML starting with the initial estimator.

We provided asymptotic results and extensive Monte Carlo results showing that the final estimators are highly efficient and maintain the same robustness level as the initial ones.

Acknowledgements

All statistical analysis were performed on SCSCF (www.dais.unive.it/scscf), a multiprocessor cluster system owned by Ca' Foscari University of Venice running under GNU/Linux.

This research was partially supported by the Italian-Argentinian project “Metodi robusti per la previsione del costo e della durata della degenza ospedaliera” funded by the collaboration program MINCYT-MAE AR14MO6.

Víctor Yohai research was also partially supported by Grants 20020130100279 from Universidad of Buenos Aires, PIP 112-2008-01-00216 and 112-2011-01-00339 from CONICET and PICT 2011-0397 from ANPCYT.

A Appendix

The Appendix contains in subsection A.1 the proof for the consistency of the TQ_τ estimator for both cases: without and with covariates, in subsection A.2 the derivation of the asymptotic distribution of the 1TML and 2TML estimators and in subsection A.3 the definition of the one step weighted likelihood estimators.

A.1 Consistency of the TQ_τ estimator

Consistency and $n^{1/2}$ -consistency are proved for the TQ_τ estimator for the case without covariates and for the regression case in the next two subsections.

A.1.1 The case without covariables

Consider the TQ_τ estimator with trimming proportion α defined in Section 2.3 to estimate the parameter $\boldsymbol{\theta}$ of a GLG distribution. We need the following assumptions.

B1 Θ is a compact set.

B2 For all $\boldsymbol{\theta}$ and all $y \in \mathbb{R}$ we have $0 < F_{\boldsymbol{\theta}}(y) < 1$.

B3 $F_{\boldsymbol{\theta}}(y)$ has a continuous and bounded density $f_{\boldsymbol{\theta}}(y) > 0$ in y and $\boldsymbol{\theta}$.

B4 Given $\boldsymbol{\theta}_1 \neq \boldsymbol{\theta}_2$ there is only a finite number of values y such that $F_{\boldsymbol{\theta}_1}(y) = F_{\boldsymbol{\theta}_2}(y)$.

B5 Let $Q(u, \boldsymbol{\theta})$ be defined as the unique value q such that $F_{\boldsymbol{\theta}}(q) = u$.

Note that **B1-B3** imply that $Q(u, \boldsymbol{\theta})$ is continuous in u and $\boldsymbol{\theta}$ and strictly increasing in u . Put $\Delta Q(u, \boldsymbol{\theta}) = \partial Q(u, \boldsymbol{\theta}) / \partial \boldsymbol{\theta}$ and $E_0 = \{\mathbf{e} \in \mathbb{R}^p : \|\mathbf{e}\| = 1\}$. Then, for all $\mathbf{e} \in E_0$, there is only a finite number of values u such that $\Delta Q(u, \boldsymbol{\theta}_0)^\top \mathbf{e} = 0$.

Theorem 1 Assume ρ_1 and ρ_2 satisfy assumption **A1** and **B1-B5**. Let $\tilde{\boldsymbol{\theta}}$ be the TQ_τ estimator. Then,

$$\tilde{\boldsymbol{\theta}} \rightarrow \boldsymbol{\theta} \quad a.s., n \rightarrow \infty$$

and

$$n^{1/2} \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) = O_P(1).$$

Proof. The proof of consistency is similar to the proof of Theorem 1 in Agostinelli et al. (2014b) while the proof of $n^{1/2}$ consistency is similar to the proof of Theorem 2 in Agostinelli et al. (2014b). The only difference is that we replace the empirical distribution F_n by the Kaplan-Meier distribution \tilde{F}_n . Note that the only property of F_n that was used in the proof of Theorem 2 is

$$n^{1/2} \sup_t |F_n^{-1}(t) - F_0^{-1}(t)| = O_p(1).$$

Then, in order to prove this Theorem, it is enough to prove that

$$n^{1/2} \sup_t \left| \tilde{F}_n^{-1}(t) - F_0^{-1}(t) \right| = O_p(1), \quad (38)$$

and this is proved in Lemma 2 below. Since we are considering an α -trimmed $Q\tau$, we also need a slightly modified version of Lemma 2 in Agostinelli et al. (2014b), where instead of considering the range of quantiles in the interval $[0.1, 0.9]$, we consider the range $[0.10, 1 - \alpha]$. The proof is exactly the same and it is omitted. ■

Suppose now that F_n is a sequence of estimators of a distribution function F_0 with support (a, b) , where a may be $-\infty$ and b may be $+\infty$. We consider the random variables $V_n = n^{1/2} \sup_x |F_n(x) - F_0(x)|$ and $S_n = n^{1/2} \sup_{\alpha \leq u \leq \beta} |F_n^{-1}(u) - F_0^{-1}(u)|$, where $0 < \alpha < \beta < 1$. The following assumptions are required.

C1 V_n is bounded in probability.

C2 $f_0(x) = F'_0(x)$ is continuous and positive in (a, b) .

Lemma 1 Assume **C1** and **C2**. Then, S_n is bounded in probability.

Proof. Take $\delta = \min(F_0^{-1}(\alpha)/2, (1 - F_0^{-1}(\beta))/2)$ and let

$$\eta = \inf_{F^{-1}(\alpha) - \delta \leq x \leq F^{-1}(\beta) + \delta} f_0(x). \quad (39)$$

By **C1**, for every $\varepsilon > 0$ there exists K_1 such

$$P(V_n > K_1) < \varepsilon. \quad (40)$$

Take n_0 such that

$$K_1/n_0^{1/2} < \delta \quad (41)$$

and K_0 , such that

$$K_0 > K_1/\eta. \quad (42)$$

We will show that, for $n \geq n_0$,

$$P(S_n > K_0) < \varepsilon.$$

To prove this, it is enough to show that for $n \geq n_0$

$$\{S_n > K_0\} \subset \{V_n > K_1\}. \quad (43)$$

By (40), to prove this it is enough to show that, for $n \geq n_0$ and for all u such that $\alpha \leq u \leq \beta$, we have

$$V_n \leq K_1 \implies n^{1/2}(F_n^{-1}(u) - F_0^{-1}(u)) \leq K_0 \quad (44)$$

and

$$V_n \leq K_1 \implies n^{1/2}(F_n^{-1}(u) - F_0^{-1}(u)) \geq -K_0. \quad (45)$$

Suppose that $n \geq n_0$, $V_n \leq K_1$ and $\alpha \leq u \leq \beta$. Then, using (39), (41), (42), we get

$$\begin{aligned} F_n \left(F_0^{-1}(u) + \frac{K_0}{n^{1/2}} \right) &> F_0 \left(F_0^{-1}(u) + \frac{K_0}{n^{1/2}} \right) - \frac{K_1}{n^{1/2}} \\ &> F_0(F_0^{-1}(u)) + \frac{\eta K_0}{n^{1/2}} - \frac{K_1}{n^{1/2}} > u + \frac{\eta K_0 - K_1}{n^{1/2}} > u, \end{aligned}$$

and therefore $F_n^{-1}(u) \leq F_0^{-1}(u) + K_0/n^{1/2}$. Then (44) holds. The proof of (45) is similar. ■

Lemma 2 Assume ρ_1 and ρ_2 satisfy assumption **A1** and **B1-B5** and **C1-C2** then

$$n^{1/2} \sup_t \left| \tilde{F}_n^{-1}(t) - F_0^{-1}(t) \right| = O_p(1).$$

Proof. Let \tilde{F}_n be the Kaplan-Meier distribution applied to $Y_n^* = (y_1^*, \dots, y_n^*)$ and Δ_n . Breslow and Crowley (1974) showed that

$$n^{1/2} \sup_t \left| \tilde{F}_n(t) - F_0(t) \right| = O_p(1),$$

and hence, by Lemma 1, we have the result. ■

A.1.2 The case with covariables

We now move to the regression estimators defined in Section 4. We suppose that $(y_i^*, \mathbf{x}_i, \delta_i)$ ($1 \leq i \leq n$, $\mathbf{x}_i \in \mathbb{R}^p$) is a sample of observations which follow the model (29), i.e.,

$$y_i = \mu_0 + \boldsymbol{\beta}_0^\top \mathbf{x}_i + \sigma_0 u_i, \quad i = 1, \dots, n,$$

and $u_i \sim f_{0,1,\lambda_0}$. Let $\bar{\boldsymbol{\beta}}_n$ be an estimator of $\boldsymbol{\beta}_0$. We consider the following assumptions

D1 $n^{1/2}(\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)$ is bounded in probability.

D2 There exist C such that for all i we have $\|\mathbf{x}_i\| \leq C$.

We define

$$\begin{aligned} v_i &= y_i - \boldsymbol{\beta}_0^\top \mathbf{x}_i = \mu_0 + \sigma_0 u_i, \\ v_i^* &= y_i^* - \boldsymbol{\beta}_0^\top \mathbf{x}_i = \min(v_i, c_i - \boldsymbol{\beta}_0^\top \mathbf{x}_i), \end{aligned}$$

and the residuals

$$\omega_i^* = y_i^* - \bar{\boldsymbol{\beta}}_n^\top \mathbf{x}_i.$$

Then,

$$\omega_i^* = \min(v_i, c_i - \boldsymbol{\beta}_0^\top \mathbf{x}_i) + (\boldsymbol{\beta}_0 - \bar{\boldsymbol{\beta}}_n)^\top \mathbf{x}_i = v_i^* + (\bar{\boldsymbol{\beta}}_n - \boldsymbol{\beta}_0)^\top \mathbf{x}_i.$$

We consider that the estimate $\tilde{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$ is the TQ τ based on the residuals (ω_i^*, δ_i) , $1 \leq i \leq n$ and we let $\tilde{\boldsymbol{\gamma}} = (\bar{\boldsymbol{\beta}}, \tilde{\boldsymbol{\theta}})$ and $\boldsymbol{\gamma}_0 = (\boldsymbol{\beta}_0, \boldsymbol{\theta}_0)$. Suppose that, using a sample $X_n = (x_1, \dots, x_n)$ and the censorship indicators $\Delta_n = (\delta_1, \dots, \delta_n)$, we have an estimator $F_n^{X_n, \Delta_n}(x)$ of a distribution F_0 . We assume that

E1 $n^{1/2} \sup_x |F_n^{X_n, \Delta_n}(x) - F_0(x)|$ is bounded in probability.

E2 Suppose that given two samples $X_n = (x_1, \dots, x_n)$ and $X_n^* = (x_1^*, \dots, x_n^*)$ such that $x_i \geq x_i^*$ for all i , then $F_n^{X_n, \Delta_n}(x) \leq F_n^{X_n^*, \Delta_n}(x)$.

E3 Given a real number z and a sample $X_n = (x_1, \dots, x_n)$, let $X_n + z = (x_1 + z, \dots, x_n + z)$. Then $F_n^{X_n + z, \Delta_n}(x) = F_n^{X_n, \Delta_n}(x - z)$.

E4 Let $Q_n = (q_1, \dots, q_n)$ be a sequence of random variables such that $n^{1/2} \sup_{1 \leq i \leq n} |q_i|$ is bounded in probability.

Theorem 2 Assume **D1-D2** and **E1-E4** then $\tilde{\gamma}$ is a consistent estimator of γ_0 and furthermore

$$n^{1/2} \sup_t |\tilde{\gamma} - \gamma_0| = O_p(1).$$

In order to prove Theorem 2 we need the following Lemma 3.

Lemma 3 Assume **E1-E4**. Then, $U_n = n^{1/2} \sup_x |F_n^{X_n+Q_n, \Delta_n}(x) - F_0(x)|$ is bounded in probability.

Proof. We have to prove that given $\varepsilon > 0$ there exists K_0 such that

$$P(\sup_n U_n > K_0) < \varepsilon. \quad (46)$$

Let $V_n = n^{1/2} \sup_x |F_n^{X_n, \Delta_n}(x) - F_0(x)|$. By **E1**, there exists K_1 such that

$$P(\sup_n V_n > K_1) < \varepsilon/3. \quad (47)$$

Let $W_n = n^{1/2} \sup_{1 \leq i \leq n} |q_i|$. By **E4**, there exists K_2 such that

$$P(\sup_n W_n > K_2) < \varepsilon/3. \quad (48)$$

Let $K_3 = \sup_x f_0(x)$ and put $K_0 = K_1 + K_2 K_3$. Then, by **E2** and **E3** we have

$$\begin{aligned} P(U_n > K_0) &= P\left(n^{1/2} \sup_x |F_n^{X_n+Q_n, \Delta_n}(x) - F_0(x)| > K_0\right) \\ &\leq P\left(n^{1/2} \sup_x \left|F_n^{X_n + \frac{K_2}{n^{1/2}}, \Delta_n}(x) - F_0(x)\right| > K_0\right) \\ &\quad + P\left(n^{1/2} \sup_x \left|F_n^{X_n - \frac{K_2}{n^{1/2}}, \Delta_n}(x) - F_0(x)\right| > K_0\right) + P(W_n > K_2). \end{aligned}$$

By **E3** we have

$$n^{1/2} \sup_x \left|F_0\left(x - \frac{K_2}{n^{1/2}}\right) - F_0(x)\right| \leq K_2 K_3. \quad (49)$$

Then, using (49), we get

$$\begin{aligned}
& P \left(n^{1/2} \sup_x \left| F_n^{X_n + \frac{K_2}{n^{1/2}}, \Delta_n}(x) - F_0(x) \right| > K_0 \right) \\
&= P \left(n^{1/2} \sup_x \left| F_n^{X_n, \Delta_n} \left(x - \frac{K_2}{n^{1/2}} \right) - F_0(x) \right| > K_1 + K_2 K_3 \right) \\
&\leq P \left(n^{1/2} \sup_x \left| F_n^{X_n, \Delta_n} \left(x - \frac{K_2}{n^{1/2}} \right) - F_0 \left(x - \frac{K_2}{n^{1/2}} \right) \right| \right. \\
&\quad \left. + n^{1/2} \sup_x \left| F_0 \left(x - \frac{K_2}{n^{1/2}} \right) - F_0(x) \right| > K_1 + K_2 K_3 \right) \\
&\leq P \left(n^{1/2} \sup_x \left| F_n^{X_n, \Delta_n} \left(x - \frac{K_2}{n^{1/2}} \right) - F_0 \left(x - \frac{K_2}{n^{1/2}} \right) \right| > K_1 \right) \\
&\leq P(V_n > K_1) \leq \frac{\varepsilon}{3}.
\end{aligned} \tag{50}$$

Similarly we can prove that

$$P \left(n^{1/2} \sup_x \left| F_n^{X_n - \frac{K_2}{n^{1/2}}, \Delta_n}(x) - F_0(x) \right| > K_0 \right) \leq \frac{\varepsilon}{3}. \tag{51}$$

Then, from (48), (50) and (51) we get (46). ■

Proof of Theorem 2. Let F_n be the Kaplan-Meier distribution applied to $V_n^* = (v_1^*, \dots, v_n^*)$ and Δ_n , let \tilde{F}_n be the Kaplan-Meier distribution applied to (Ω_n^*, Δ_n) , where $\Omega_n^* = (\omega_1^*, \dots, \omega_n^*)$ and $\Delta_n = (\delta_1, \dots, \delta_n)$ and let F_0 be the distribution of the v_i s. The proof is similar to that of Theorem 1 and hence it is sufficient to prove that

$$n^{1/2} \sup_t |\tilde{F}_n^{-1}(t) - F_0^{-1}(t)| = O_P(1).$$

Breslow and Crowley (1974) showed that

$$n^{1/2} \sup_t |F_n(t) - F_0(t)| = O_P(1).$$

Then, since the Kaplan-Meier distribution satisfies **E2** and **E3**, Lemma 3 implies

$$n^{1/2} \sup_t |\tilde{F}_n(t) - F_0(t)| = O_P(1).$$

On the other hand $\Omega_n^* = V_n^* + Q_n$ where $Q_n = (\beta_0 - \bar{\beta}_n)^T \mathbf{x}_i$ satisfies **E4** and by Lemma 1 the result holds. ■

A.2 Asymptotic distribution of the 1TML estimator

In this Section we study the asymptotic behavior of the 1TML estimator. Note that the asymptotic behavior of the 2TML is obtained in a very similar way and hence it is not reported. We suppose that $\tilde{\gamma}$ is an initial $n^{1/2}$ consistent estimator, e.g. the TQ τ estimator, and that γ_0 is the true parameter value. We need the following assumptions:

F1 $n^{1/2}(\tilde{\gamma} - \gamma_0)$ is bounded in probability.

F2 The matrix $\mathbf{G}(\gamma) = E(\nabla_{\gamma} \mathbf{v}(\mathbf{x}, y, \delta, \gamma))$ (where $\mathbf{v}(\mathbf{x}, y, \delta, \gamma)$ is defined in equation (30)) is non singular.

F3 The vector \mathbf{x} has second order moments.

Theorem 3 Assume (i) the weight function w satisfies **A2**, (ii) the function w is continuously differentiable (iii) **F1-F3** holds, then the 1TML estimator $\hat{\gamma}_1$ is such that

$$\sqrt{n}(\hat{\gamma}_1 - \gamma_0) \xrightarrow{\mathcal{L}} N(0, \mathbf{G}(\gamma_0)^{-1} \mathbf{M}(\gamma_0) \mathbf{G}(\gamma_0)^{-\top}).$$

To prove the Theorem we need some additional notations and the lemma stated below. Let

$$\boldsymbol{\eta}(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*) = (\eta_1(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*), \dots, \eta_{p+3}(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*))^\top$$

so that, for given γ_1 and γ_2 ,

$$\eta_k(\mathbf{z}_i, \gamma_1, \gamma_2, \vartheta^*) = E_{\gamma_1} [w(\mathbf{z}, \gamma_2, \vartheta^*) v_k(\mathbf{z}, \gamma_1) | \mathbf{z}_i] \quad k = 1, \dots, (p+3),$$

where v_1, \dots, v_{p+3} are the score functions evaluated at γ_1 . Then, the functions $u_k(\gamma, \tilde{\gamma}, \vartheta^*)$ in (33) can be written as follows:

$$u_k(\gamma, \tilde{\gamma}, \vartheta^*) = \frac{1}{n} \sum_{i=1}^n \eta_k(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*), \quad k = 1, \dots, p+3.$$

In addition, for given $\gamma_1, \gamma_2, \gamma_3$, let

$$\mathbf{h}(\gamma_1, \gamma_2, \gamma_3, \vartheta^*) = (h_1(\gamma_1, \gamma_2, \gamma_3, \vartheta^*), \dots, h_{p+3}(\gamma_1, \gamma_2, \gamma_3, \vartheta^*))^\top,$$

where

$$h_k(\gamma_1, \gamma_2, \gamma_3, \vartheta^*) = E_{\gamma_1} [\eta_k(\mathbf{z}, \gamma_2, \gamma_3, \vartheta^*)], \quad k = 1, \dots, p+3.$$

We also have:

$$h_k(\gamma, \tilde{\gamma}, \gamma, \vartheta^*) = E_\gamma [w(\mathbf{z}, \tilde{\gamma}, \vartheta^*) v_k(\mathbf{z}, \gamma)], \quad k = 1, \dots, p+3.$$

Therefore, equations (33) can be written as

$$\frac{1}{n} \sum_{i=1}^n \boldsymbol{\eta}(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*) = \mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*).$$

Finally, notice that

$$\mathbf{J}(\tilde{\gamma}, \vartheta^*) = \frac{1}{n} \sum_{i=1}^n \nabla_\gamma \boldsymbol{\eta}(\mathbf{z}_i, \gamma, \tilde{\gamma}, \vartheta^*)|_{\gamma=\tilde{\gamma}}.$$

Lemma 4 *Under the assumptions of Theorem 3, we have*

$$\sqrt{n}|\mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\gamma_0, \tilde{\gamma}, \gamma_0, \vartheta^*)| \xrightarrow{P} \mathbf{0} \quad n \rightarrow \infty.$$

Proof. We prove the lemma by showing that

$$\sqrt{n}|\mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\gamma_0, \gamma_0, \gamma_0, \vartheta^*)| \xrightarrow{P} \mathbf{0} \quad n \rightarrow \infty, \quad (52)$$

and

$$\sqrt{n}|\mathbf{h}(\gamma_0, \gamma_0, \gamma_0, \vartheta^*) - \mathbf{h}(\gamma_0, \tilde{\gamma}, \gamma_0, \vartheta^*)| \xrightarrow{P} \mathbf{0} \quad n \rightarrow \infty. \quad (53)$$

Let us consider (52). We can write

$$\sqrt{n}(\mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\gamma_0, \gamma_0, \gamma_0, \vartheta^*)) = \nabla_\gamma \mathbf{h}(\gamma, \gamma, \gamma, \vartheta^*)|_{\gamma=\gamma^*} \sqrt{n}(\tilde{\gamma} - \gamma),$$

where γ^* is between $\tilde{\gamma}$ and γ_0 . According to **F1**, the second factor is bounded in probability. The first factor is asymptotically zero since

$$\begin{aligned} \lim_{n \rightarrow \infty} \nabla_\gamma \mathbf{h}(\gamma, \gamma, \gamma, \vartheta^*)|_{\gamma=\gamma^*} &= \nabla_\gamma \lim_{n \rightarrow \infty} \mathbf{h}(\gamma, \gamma, \gamma, \vartheta^*)|_{\gamma=\gamma^*} \\ &= \nabla_\gamma \mathbf{h}(\gamma, \gamma, \gamma, 0)|_{\gamma=\gamma_0} \\ &= \nabla_\gamma E_\gamma(\mathbf{v}(\mathbf{z}, \gamma))|_{\gamma=\gamma_0} \end{aligned}$$

but $E_\gamma(\mathbf{v}(\mathbf{z}, \gamma)) = \mathbf{0}$ for all γ . For (53) we can write

$$\begin{aligned} &\sqrt{n}(\mathbf{h}(\gamma_0, \gamma_0, \gamma_0, \vartheta^*) - \mathbf{h}(\gamma_0, \tilde{\gamma}, \gamma_0, \vartheta^*)) \\ &= \sqrt{n}E_{\gamma_0}(\mathbf{v}(\mathbf{z}, \gamma_0)(w(\mathbf{z}, \tilde{\gamma}, \vartheta^*) - w(\mathbf{z}, \gamma_0, \vartheta^*))) \\ &= E_{\gamma_0}(\mathbf{v}(\mathbf{z}, \gamma_0) \nabla_\gamma w(\mathbf{z}, \gamma, \vartheta^*)|_{\gamma=\gamma^*} \sqrt{n}(\tilde{\gamma} - \gamma)), \end{aligned}$$

where γ^* is between $\tilde{\gamma}$ and γ_0 . The second term is such that

$$\begin{aligned}\lim_{n \rightarrow \infty} \nabla_{\gamma} w(\mathbf{z}, \gamma, \vartheta^*)|_{\gamma=\gamma^*} &= \nabla_{\gamma} \lim_{n \rightarrow \infty} w(\mathbf{z}, \gamma, \vartheta^*)|_{\gamma=\gamma^*} \\ &= \nabla_{\gamma} w(\mathbf{z}, \gamma, 0)|_{\gamma=\gamma_0},\end{aligned}$$

but $w(\mathbf{z}, \gamma, 0) \equiv 1$ for all γ and hence $\nabla_{\gamma} w(\mathbf{z}, \gamma, 0) \equiv 0$ for all γ . ■

Lemma 5 *Let \mathbf{z}_i , $1 \leq i \leq n$, be i.i.d. random vectors in \mathbb{R}^q with distribution G and $\boldsymbol{\zeta}$ a parameter in \mathbb{R}^s . Let $f(\mathbf{z}, \boldsymbol{\zeta}) : \mathbb{R}^q \mathbb{R}^s \rightarrow \mathbb{R}$ be a function continuously differentiable with respect to $\boldsymbol{\zeta}$, such that (i) $E_G(f(\mathbf{z}, \boldsymbol{\zeta})) = 0$ for all $\boldsymbol{\zeta}$, (ii) $E_G(f^2(\mathbf{z}, \boldsymbol{\zeta}_0)) < \infty$, and (iii) there exist $\delta > 0$ and $c(\mathbf{z}) : \mathbb{R}^q \rightarrow \mathbb{R}$ satisfying*

$$\sup_{\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_0\| \leq \delta} \left| \frac{\partial f(\mathbf{z}, \boldsymbol{\zeta})}{\partial \boldsymbol{\zeta}} \right| \leq c(\mathbf{z}),$$

where $E_G(c^2(\mathbf{z})) < \infty$. Let $\boldsymbol{\zeta}_n$ be a sequence of random variables converging to $\boldsymbol{\zeta}_0$ a.s.. Then,

$$n^{-1/2} \sum_{i=1}^n (f(\mathbf{z}_i, \boldsymbol{\zeta}_n) - f(\mathbf{z}_i, \boldsymbol{\zeta}_0)) \xrightarrow{P} 0. \quad (54)$$

Proof. Take $\varepsilon > 0$ and consider the sequence of processes $S_n(\boldsymbol{\zeta})$, $\boldsymbol{\zeta} \in S = \{\boldsymbol{\zeta} : \|\boldsymbol{\zeta} - \boldsymbol{\zeta}_0\| \leq \varepsilon\}$ in the space C of the continuous functions in S :

$$S_n(\boldsymbol{\zeta}) = \frac{1}{n^{1/2}} \sum_{i=1}^n f(\mathbf{z}_i, \boldsymbol{\zeta}).$$

To prove the Lemma is enough to show that $S_n(\boldsymbol{\zeta})$ is tight. According to Theorem 1 of Jain and Marcus (1975), $S_n(\boldsymbol{\zeta})$ converges in distribution to a Gaussian process $S(\lambda)$ and therefore it is tight. ■ Finally, we report Lemma A3.1 of Marazzi et al. (2009)

Lemma 6 (Lemma A3.1 in Marazzi et al. (2009)) *Let \mathbf{z}_i , $1 \leq i \leq n$, be i.i.d. random vectors in \mathbb{R}^q with distribution G and $\boldsymbol{\zeta}$ a parameter in \mathbb{R}^s . Let $f(\mathbf{z}, \boldsymbol{\zeta}) : \mathbb{R}^q \mathbb{R}^s \rightarrow \mathbb{R}^k$ be a continuous function, such that there exist $\delta > 0$ and $c(\mathbf{z}) : \mathbb{R}^q \rightarrow \mathbb{R}$ satisfying*

$$\sup_{\|\boldsymbol{\zeta} - \boldsymbol{\zeta}_0\| \leq \delta} |f(\mathbf{z}, \boldsymbol{\zeta})| \leq c(\mathbf{z}) \quad a.s.,$$

and $E_G(c(\mathbf{z})) < \infty$. Let $\boldsymbol{\zeta}_n$ be a sequence of random variables converging to $\boldsymbol{\zeta}_0$ in probability. Then,

$$\frac{1}{n} \sum_{i=1}^n f(\mathbf{z}_i, \boldsymbol{\zeta}_n) \xrightarrow{P} E_G(f(\mathbf{z}_i, \boldsymbol{\zeta}_0)).$$

Proof of Theorem 3. We recall that the 1TML estimator is defined as

$$\hat{\gamma}_1 = \tilde{\gamma} - J(\tilde{\gamma}, \vartheta^*)^{-1} \frac{1}{n} \sum_{i=1}^n (\boldsymbol{\eta}(\mathbf{z}_i, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*))$$

We consider an expansion of $\boldsymbol{\eta}$ in its second argument around γ_0 has follows

$$\sum_{i=1}^n \boldsymbol{\eta}(\mathbf{z}_i, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*) = \sum_{i=1}^n \boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) + J(\gamma^*, \vartheta^*)(\tilde{\gamma} - \gamma_0)$$

and γ^* is between $\tilde{\gamma}$ and γ_0 . Then,

$$\begin{aligned} \sqrt{n}(\hat{\gamma}_1 - \gamma_0) &= \sqrt{n}(\tilde{\gamma} - \gamma_0) \\ &\quad - nJ(\tilde{\gamma}, \vartheta^*)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n [\boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*)] \quad (55) \\ &\quad - \sqrt{n}J(\tilde{\gamma}, \vartheta^*)^{-1} J(\gamma^*, \vartheta^*)(\tilde{\gamma} - \gamma_0) \\ &= -nJ(\tilde{\gamma}, \vartheta^*)^{-1} \frac{1}{\sqrt{n}} \sum_{i=1}^n [\boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*)] \\ &\quad + (1 - J(\tilde{\gamma}, \vartheta^*)^{-1} J(\gamma^*, \vartheta^*)) \sqrt{n}(\tilde{\gamma} - \gamma_0). \end{aligned}$$

By Lemma 4, we have

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n [\boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\tilde{\gamma}, \tilde{\gamma}, \tilde{\gamma}, \vartheta^*)] \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n [\boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\gamma_0, \tilde{\gamma}, \gamma_0, \vartheta^*)] + o_p(1). \end{aligned}$$

Let $\Sigma(\gamma_0)$ be the asymptotic covariance matrix of the maximum likelihood estimators and by Lemma 5

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n [\boldsymbol{\eta}(\mathbf{z}_i, \gamma_0, \tilde{\gamma}, \vartheta^*) - \mathbf{h}(\gamma_0, \tilde{\gamma}, \gamma_0, \vartheta^*)] \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Sigma(\gamma_0)). \quad (56)$$

Finally, by Lemma 6 we have

$$\mathbf{J}(\tilde{\gamma}, \vartheta^*) = \mathbf{G}(\gamma_0) + o_P(1) \quad (57)$$

and $\mathbf{J}(\gamma^*, \vartheta^*) = \mathbf{G}(\gamma_0) + o_P(1)$. Hence, $(1 - \mathbf{J}(\tilde{\gamma}, \vartheta^*)^{-1} \mathbf{J}(\gamma^*, \vartheta^*)) \xrightarrow{P} 0$ by Slutsky Theorem. The asymptotic normality of the 1TML estimator follows from (55), (56), (57) and the Slutsky theorem. ■ We now consider the case without covariates. Next Theorem follows immediately from Theorem 3.

Theorem 4 *Let us consider observations (y_i^*, δ_i) ($1 \leq i \leq n$) following a GLG model with parameter $\boldsymbol{\theta}_0 = (\mu_0, \sigma_0, \lambda_0)$. Let $\hat{\boldsymbol{\theta}}$ be an initial estimator. Assume: (i) $n^{1/2}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)$ is bounded in probability; (ii) the matrix $\mathbf{G}(\boldsymbol{\theta})$ is non singular; (iii) the weight function w satisfies **A2**; (iv) the function w is continuously differentiable. Then, the 1TML estimator $\hat{\boldsymbol{\theta}}_1$ is such that:*

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}_0) \xrightarrow{\mathcal{L}} N(0, \mathbf{G}(\boldsymbol{\theta}_0)^{-1} \mathbf{M}(\boldsymbol{\theta}_0) \mathbf{G}(\boldsymbol{\theta}_0)^{-\top})$$

We recall that $\mathbf{M}(\boldsymbol{\theta}) = E(\mathbf{v}(y, \delta, \boldsymbol{\theta}) \mathbf{v}(y, \delta, \boldsymbol{\theta})^\top)$ and $\mathbf{G}(\boldsymbol{\theta}) = E(\nabla_{\boldsymbol{\theta}} \mathbf{v}(y, \delta, \boldsymbol{\theta}))$ and $\mathbf{v}(y, \delta, \boldsymbol{\theta})$ is defined in equation (11).

Remark 1 *In practice, especially when the sample size is not very large, and ϑ^* is far from the asymptotic value 0, a better approximation of the covariance matrix of the 1TML estimator can be based on the following sandwich formula. Let*

$$\begin{aligned} \Lambda(\mathbf{z}, \hat{\gamma}_1, \vartheta^*) \\ = (\boldsymbol{\eta}(\mathbf{z}, \hat{\gamma}_1, \hat{\gamma}_1, \vartheta^*) - \mathbf{h}(\hat{\gamma}_1, \hat{\gamma}_1, \hat{\gamma}_1, \vartheta^*))(\boldsymbol{\eta}(\mathbf{z}, \hat{\gamma}_1, \hat{\gamma}_1, \vartheta^*) - \mathbf{h}(\hat{\gamma}_1, \hat{\gamma}_1, \hat{\gamma}_1, \vartheta^*))^\top. \end{aligned}$$

Then, the covariance matrix $\boldsymbol{\Sigma}(\hat{\gamma}_1)$ of the 1TML estimator can be estimated by

$$\hat{\boldsymbol{\Sigma}}(\hat{\gamma}_1) = \mathbf{J}(\hat{\gamma}_1, \vartheta^*)^{-1} \left(\frac{1}{n} \sum_{i=1}^n \Lambda(\mathbf{z}, \hat{\gamma}_1, \vartheta^*) \right) \mathbf{J}(\hat{\gamma}_1, \vartheta^*)^{-\top}.$$

Remark 2 *For the case without covariates*

$$\mathbf{G}(\boldsymbol{\theta}) = E(\delta \mathbf{G}_d + (1 - \delta) \mathbf{G}_s),$$

where the matrix \mathbf{G}_d corresponds to the uncensored observations and \mathbf{G}_s to the censored observations. Hence, the elements of \mathbf{G}_d are

$$\begin{aligned}
g_{d11} &= \frac{\partial d_1}{\partial \mu} = -\frac{1}{\sigma^2} \xi'_\lambda(u), \\
g_{d12} &= \frac{\partial d_1}{\partial \sigma} = -\frac{1}{\sigma^2} (\xi_\lambda(u) + \xi'_\lambda(u)u), \\
g_{d13} &= \frac{\partial d_1}{\partial \lambda} = \frac{1}{\sigma} \dot{\xi}_\lambda(u), \\
\\
g_{d21} &= \frac{\partial d_2}{\partial \mu} = -\frac{1}{\sigma^2} (\xi'_\lambda(u)u + \xi_\lambda(u)), \\
g_{d22} &= \frac{\partial d_2}{\partial \sigma} = -\frac{1}{\sigma^2} (2\xi_\lambda(u)u + \xi'_\lambda(u)u^2 + 1), \\
g_{d23} &= \frac{\partial d_2}{\partial \lambda} = \frac{1}{\sigma} \dot{\xi}_\lambda(u)u, \\
\\
g_{d31} &= \frac{\partial z_3}{\partial \mu} = -\frac{1}{\sigma} \psi'_\lambda(u), \\
g_{d32} &= \frac{\partial z_3}{\partial \sigma} = -\frac{1}{\sigma} \psi'_\lambda(u)u, \\
g_{d33} &= \frac{\partial z_3}{\partial \lambda} = \dot{\psi}_\lambda(u).
\end{aligned}$$

A.3 The weighted likelihood estimator

In the next two subsections we extend the results in Agostinelli et al. (2014a) concerning the 1SWL estimator to the case with censored observations and to the regression case.

A.3.1 The case without covariables

Markatou et al. (1998) introduced the weighted likelihood estimators for the case of non censored data in continuous models. Here, we extend this estimator to the case of censored data. Assume that $w(y, \boldsymbol{\theta})$ is a given weight function and that an initial highly robust and consistent but not necessarily efficient estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}_0$ – e.g., a TQ τ estimator $\hat{\boldsymbol{\theta}}$ defined in Section 2.3

– is available. Then, we define the *weighted likelihood (WL) estimator* for censored observations as a solution of the equation

$$\mathbf{u}(\mathbf{y}^*, \boldsymbol{\delta}, \boldsymbol{\theta}) = \sum_{i=1}^n w(y_i^*, \boldsymbol{\theta}) \mathbf{v}(y_i^*, \delta_i, \boldsymbol{\theta}) = \mathbf{0}, \quad (58)$$

where $\mathbf{v}(y_i^*, \delta_i, \boldsymbol{\theta})$ is the same as in (12). When $w(y_i^*, \boldsymbol{\theta}) \equiv 1$, (58) coincides with (12). Following Markatou et al. (1998) we define the weight function as

$$w(y, \boldsymbol{\theta}) = \min \left(1, \frac{[A(r_p(y, \boldsymbol{\theta})) + 1]^+}{r_p(y, \boldsymbol{\theta}) + 1} \right), \quad (59)$$

where $[x]^+ = \max(0, x)$, $r_p(y, \boldsymbol{\theta})$ is a convenient variant of *Pearson residual*, measuring the agreement between the data and the assumed model, and the function $A(\cdot)$ is a *residual adjustment function* (Lindsay, 1994).

In order to define the Pearson residual, we proceed as follows. Let $H_{n,\boldsymbol{\theta}}$ be the semiparametric cdf (13) and $f_{n,\boldsymbol{\theta}}^*(y) = \int k(y, t, h) dH_{n,\boldsymbol{\theta}}$ be a semi-parametric kernel density estimator of $f_{\boldsymbol{\theta}}(y)$ with bandwidth h . ($f_{n,\boldsymbol{\theta}}^*$ can be approximately computed by fitting a kernel density estimator to k quantiles of $H_{n,\boldsymbol{\theta}}$, e.g., $k = 1000$.) Let $f_{\boldsymbol{\theta}}^*(y) = \int k(y, t, h) f_{\boldsymbol{\theta}}(t) dt$ be the corresponding smoothed model density. Then, the Pearson residuals are defined by

$$r_p(y, \boldsymbol{\theta}) = [f_{n,\boldsymbol{\theta}}^*(y) - f_{\boldsymbol{\theta}}^*(y)] / f_{\boldsymbol{\theta}}^*(y).$$

Equation (58) can be solved using an iterative algorithm as in Markatou et al. (1998).

Following Agostinelli and Markatou (1998) and Agostinelli et al. (2014a), we obtain the *one-step weighted likelihood (1SWL) estimator* for censored observations by applying one Newton-Raphson iteration to equation (58). This estimator turns out to be

$$\hat{\boldsymbol{\theta}} = \tilde{\boldsymbol{\theta}} - \mathbf{J}^{-1} \mathbf{u}(\mathbf{y}^*, \boldsymbol{\delta}, \tilde{\boldsymbol{\theta}}), \quad (60)$$

where $\mathbf{J} = \sum_{j=1}^n w(y_j^*, \tilde{\boldsymbol{\theta}}) \nabla_{\boldsymbol{\theta}} \mathbf{v}(y_j^*, \delta_j, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\tilde{\boldsymbol{\theta}}}$ is the Jacobian matrix of the functions defining the estimating equations and $\nabla_{\boldsymbol{\theta}}$ denotes differentiation with respect to $\boldsymbol{\theta}$.

A.3.2 The case with covariables

For the model (29) and parameters $\gamma = (\boldsymbol{\theta}, \boldsymbol{\beta})$, we define the fully iterated WL estimator as the solution of the equations

$$\mathbf{u}(X, \mathbf{y}^*, \boldsymbol{\delta}, \gamma) = \sum_{j=1}^n \check{w}(\mathbf{x}_j, y_j^*, \gamma) \mathbf{v}(\mathbf{x}_j, y_j^*, \delta_j, \gamma) = \mathbf{0},$$

where

$$\check{w}(\mathbf{x}_j, y_j^*, \gamma) = w(y_j^* - \mathbf{x}_j^\top \boldsymbol{\beta}, \boldsymbol{\theta}),$$

and w is given by (59). Besides, the *1SWL regression estimator* is given by

$$\hat{\gamma} = \tilde{\gamma} - \mathbf{J}^{-1} \mathbf{u}(X, \mathbf{y}^*, \boldsymbol{\delta}, \tilde{\gamma})$$

and $\mathbf{J} = \sum_{i=1}^n \check{w}(\mathbf{x}_i, y_i, \tilde{\gamma}) \nabla_{\gamma} \mathbf{v}(\mathbf{x}_i, y_i^*, \tilde{\gamma})|_{\gamma=\tilde{\gamma}}$ is a $(3+p) \times (3+p)$ Jacobian matrix, The weights are based on Pearson residuals comparing $F_{\tilde{\boldsymbol{\theta}}^*}$ to the estimator $H_{n, \tilde{\boldsymbol{\theta}}^*}$ given by (13), where the observations are $(\omega_i^*, \delta_i) = (y_i^* - \mathbf{x}_i^\top \tilde{\boldsymbol{\beta}}, \delta_i)$, $1 \leq i \leq n$.

References

- A. Abadi, F. Amanpour, C. Bajdik, and P. Yavari. Breast cancer survival analysis: Applying the generalized gamma distribution under different conditions of the proportional hazards and accelerated failure time assumptions. *International Journal of Preventive Medicine*, 3:644–651, 2012.
- C. Agostinelli and M. Markatou. A one-step robust estimator for regression based on the weighted likelihood reweighting scheme. *Statistics & Probability Letters*, 37:341–350, 1998.
- C. Agostinelli, A. Marazzi, and V.J. Yohai. Robust estimators of the generalized log-gamma distribution. *Technometrics*, 56(1):92–101, 2014a.
- C. Agostinelli, A. Marazzi, and V.J. Yohai. Supplementary material for robust estimators of the generalized log-gamma distribution, 2014b. URL <http://www.tandfonline.com/doi/suppl/10.1080/00401706.2013.818578>.

- N. Breslow and J. Crowley. A large sample study of the life table and product limit estimates under random censorship. *The Annals of Statistics*, 2:437–453, 1974.
- R.B. Fetter, Y. Shin, J.L. Freeman, R.F. Averill, and J.D. Thompson. Case mix definition by diagnosis-related groups. *Medical care*, 18(1):1–53, 1980.
- D. Gervini and V.J. Yohai. A class of robust and fully efficient regression estimators. *The Annals of Statistics*, 30(2):583–616, 2002.
- M. Greenwood. *The natural duration of cancer*. Her Majesty’s Stationery Office, London, 1926.
- P.J. Huber. *Robust statistics*. Wiley, New York, 1981.
- N. Jain and M. Marcus. Central limit theorems for $c(s)$ -valued random variables. *Journal of Functional Analysis*, 19:216–231, 1975.
- E.L. Kaplan and P. Meier. Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53(282):457–481, 1958.
- M.Y. Kim, V.G. DeGruttola, and S.W. Lagakos. Analyzing doubly censored data with covariates, with application to aids. *Biometrics*, 49:13–22, 1993.
- J.F. Lawless. Inference in the generalized gamma and log-gamma distributions. *Technometrics*, 22(3):409–419, 1980.
- B.J. Lindsay. Efficiency versus robustness: the case of minimum hellinger distance and related methods. *The Annals of Statistics*, 22:1081–1114, 1994.
- I. Locatelli and A. Marazzi. Robust parametric indirect estimates of the expected cost of a hospital stay with covariates and censored data. *Statistics in Medicine*, 32(14):2457–2466, 2013. doi: 10.1002/sim.5701.
- I. Locatelli, A. Marazzi, and Yohai V.J. Robust accelerated failure time regression. *Computational Statistics & Data Analysis*, 55:874–887, 2010.
- A. Marazzi and J.L. Muralti. RobustAFT: a package for truncated maximum likelihood fit and robust accelerated failure time regression for gaussian and log-weibull case, 2013. URL cran.r-project.org/web/packages/RobustAFT/index.html.

- A. Marazzi and V.J. Yohai. Adaptively truncated maximum likelihood regression with asymmetric errors. *Journal of Statistical Planning and Inference*, 122(1-2):271–291, 2004.
- A. Marazzi, A.J. Villar, and V.J. Yohai. Supplemental material for robust response transformations based on optimal prediction, 2009. URL www.tandfonline.com/doi/suppl/10.1198/jasa.2009.0109.
- M. Markatou, A. Basu, and B.G. Lindsay. Weighted likelihood equations with bootstrap root search. *Journal of the American Statistical Association*, 93: 740–750, 1998.
- E.M.M. Ortega, H. Bolfarine, and G.A. Paula. Influence diagnostics in generalized log-gamma regression models. *Computational Statistics & Data Analysis*, 42:165–186, 2003.
- E.M.M. Ortega, G.A. Paula, and H. Bolfarine. Deviance residuals in generalized log-gamma regression models with censored observations. *Journal of Statistical Computation and Simulation*, 78(8):747–764, 2008.
- M. Salibian-Barrera and V.J. Yohai. High breakdown point robust regression with censored data. *The Annals of Statistics*, 36:118–146, 2008.
- R.J. Serfling. *Approximation Theorems of Mathematical Statistics*. Wiley, 1980.
- G.O. Silva, E.M.M. Ortega, and V.G. Cancho. Log-weibull extended regression model: Estimation, sensitivity and residual analysis. *Statistical Methodology*, 7:614–631, 2010.
- J. Sun, Q. Liao, and M. Pagano. Regression analysis of doubly interval censored failure time data with applications to aids. *Biometrics*, 55:909–914, 1999.
- V.J. Yohai and R.H. Zamar. High breakdown estimates of regression by means of the minimization of an efficient scale. *Journal of the American Statistical Association*, 83:406–413, 1988.